

**From:** Fredman, Jeffrey  
**Sent:** Monday, October 01, 2001 3:42 PM  
**To:** STIC-Biotech/ChemLib  
**Cc:** Lacourciere, Karen  
**Subject:** FW: Sequence search approval 09/757100

I Approve.

Jeff Fredman

-----Original Message-----

**Fr m:** Lacourciere, Karen  
**Sent:** Monday, October 01, 2001 3:06 PM  
**T :** Fredman, Jeffrey  
**Subject:** Sequence search approval 09/757100

Jeff, could you approve this sequence search? Each sequence is ~20 nt's long and they are antisense targeted to one sequence. Thank-you!  
Karen

Please search the following SEQ ID NO:'s for 09/757,100 in both the commercial databases and the pending files (interference)  
3, 4, 6-12,14-18, 20, 23, 30, 31 and 33.

Please limit the length of oligos to less than 50 nucleotides long.  
Thank-you!

*Karen A. Lacourciere Ph.D.*  
✓ CM1 11D09 GAU 1635 ✓  
(703) 308-7523

**Point of Contact:**  
**Toby Port**  
**Technical Info. Specialist**  
**CM1 1E01 TEL: 308-3534**

Point of Contact:  
Toby Port  
Technical Info. Specialist  
CMI 1E01 TEL: 308-3234

P

# SEARCH REQUEST FORM

Access DB#

52048

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

## STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>1084 Port</u>	NA Sequence (#) <u>19</u>	STN _____
Searcher Phone #: <u>308-3534</u>	AA Sequence (#) _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>10/2</u>	Bibliographic _____	Dr. Link _____
Date Completed: <u>10/3</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>15</u>	Fulltext _____	Sequence Systems <u>CS</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:50 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-18  
Perfect score: 20  
Sequence: 1 tttaaccagatggtcattc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rol2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
c 1	12.8	64.0	20	9 A10280	A10280 oligonucleo
2	12.8	64.0	29	10 I82248	I82248 Sequence 32
3	12.8	64.0	29	10 I90782	I90782 Sequence 32
c 4	12.8	64.0	45	97 HUMKRT10AG	L20218 Human Kerat
c 5	12.6	63.0	35	10 I64891	I64891 Sequence 7
c 6	12.6	63.0	41	10 E13249	E13249 Oligonucleo
7	12.4	62.0	30	9 AR004965	AR004965 Sequence
8	12.4	62.0	30	10 I40158	I40158 Sequence 4

```

9 12.2 61.0 20 9 AR016116 Sequence
10 12.2 61.0 20 9 AR019114 Sequence
11 12.2 61.0 28 9 AR090906 Sequence
12 12.2 61.0 33 9 AR034456 Sequence
13 12.2 61.0 33 9 AR070401 Sequence
14 12.2 61.0 34 9 AR026654 Sequence
15 12.2 61.0 29 10 AX067989 Sequence
16 12.2 60.0 33 10 AX076373 Sequence
17 12.2 60.0 34 9 AR03829 Sequence
18 12.2 60.0 40 9 AR070497 Sequence
19 12.2 60.0 44 10 E14337 Sequence
20 11.8 59.0 16 9 A42578 Sequence
21 11.8 59.0 16 9 A42579 Sequence
22 11.8 59.0 16 9 A88767 Sequence
23 11.8 59.0 16 9 A88768 Sequence
24 11.8 59.0 23 9 A04127 Sequence
25 11.8 59.0 23 9 A067151 Sequence
26 11.8 59.0 35 9 AR075963 Sequence
27 11.8 59.0 38 9 A67648 Sequence
28 11.8 59.0 38 9 AR089824 Sequence
29 11.6 58.0 21 54 HUM66RVA Sequence
30 11.6 58.0 23 9 A37961 Sequence
31 11.6 58.0 23 10 I21766 Sequence
32 11.6 58.0 32 9 AR011102 Sequence
33 11.6 58.0 32 9 AR038261 Sequence
34 11.6 58.0 32 9 AR075191 Sequence
35 11.6 58.0 32 10 I58550 Sequence
36 11.6 58.0 32 10 I61215 Sequence
37 11.6 58.0 33 9 AR054052 Sequence
38 11.6 58.0 38 9 AR011107 Sequence
39 11.6 58.0 38 9 AR038266 Sequence
40 11.6 58.0 38 9 AR075196 Sequence
41 11.6 58.0 38 10 I58555 Sequence
42 11.6 58.0 38 10 I61220 Sequence
43 11.6 58.0 49 10 E12236 Sequence
44 11.6 58.0 49 10 E12238 Sequence
45 11.6 58.0 50 10 E22340 DNA encoding

```

## ALIGNMENTS

```

RESULT 1
A10280/c  A10280 20 bp DNA PAT 25-AUG-1993
LOCUS      oligonucleotide.
DEFINITION
ACCESSION  A10280
VERSION    A10280.1 GI:413528
KEYWORDS   synthetic construct.
SOURCE     synthetic construct.
ORGANISM   synthetic construct.
REFERENCE  1 (bases 1 to 20)
AUTHORS    Patent: DE 3802040-A 1 03-AUG-1989;
JOURNAL    Location/Qualifiers
FEATURES   source
            1..20
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT 8 a 1 c 6 g 5 t
ORIGIN

Query Match 64.0%; Score 12.8; DB 9; Length 20;
Best Local Similarity 87.5%; Pred. No. 6.3e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttcaaccagatggtc 16
Db 20 TTTCACCATGCTTC 5

RESULT 2
A10280/c  A10280 20 bp DNA PAT 25-AUG-1993
LOCUS      oligonucleotide.
DEFINITION
ACCESSION  A10280
VERSION    A10280.1 GI:413528
KEYWORDS   synthetic construct.
SOURCE     synthetic construct.
ORGANISM   synthetic construct.
REFERENCE  1 (bases 1 to 20)
AUTHORS    Patent: DE 3802040-A 1 03-AUG-1989;
JOURNAL    Location/Qualifiers
FEATURES   source
            1..20
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT 8 a 1 c 6 g 5 t
ORIGIN

Query Match 64.0%; Score 12.8; DB 9; Length 20;
Best Local Similarity 87.5%; Pred. No. 6.3e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttcaaccagatggtc 16
Db 20 TTTCACCATGCTTC 5

```

```

I82248
LOCUS      I82248 29 bp DNA PAT 10-JUN-1998
DEFINITION Sequence 32 from patent US 5712118.
ACCESSION  I82248
VERSION    I82248.1 GI:3210545
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 29)
AUTHORS    Murphy, R.F.
TITLE      Vaccine for brachamella catarrhalis
JOURNAL    Patent: US 5712118-A 32 27-JAN-1998;
FEATURES   Location/Qualifiers
            source
            1..29
            /organism="unknown"
BASE COUNT 11 a 5 c 5 g 8 t
ORIGIN

Query Match 64.0%; Score 12.8; DB 10; Length 29;
Best Local Similarity 87.5%; Pred. No. 6.3e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttcaaccagatggtc 17
Db 10 TTTCATCCAGATGATCA 25

RESULT 3
I82248 29 bp DNA PAT 01-DEC-1998
LOCUS      I82248
DEFINITION Sequence 32 from patent US 5725862.
ACCESSION  I82248
VERSION    I82248.1 GI:3935252
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 29)
AUTHORS    Murphy, R.F.
TITLE      Vaccine for brachamella catarrhalis
JOURNAL    Patent: US 5725862-A 32 10-MAR-1998;
FEATURES   Location/Qualifiers
            source
            1..29
            /organism="unknown"
BASE COUNT 11 a 5 c 5 g 8 t
ORIGIN

Query Match 64.0%; Score 12.8; DB 10; Length 29;
Best Local Similarity 87.5%; Pred. No. 6.3e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttcaaccagatggtc 17
Db 10 TTTCATCCAGATGATCA 25

RESULT 4
HUMKRT10AG/c  HUMKRT10AG 45 bp DNA PRI 06-JAN-1995
LOCUS      Human keratin 10 (KRT10) gene, partial cds including polymorphism.
DEFINITION
ACCESSION  L20218
VERSION    L20218.1 GI:307087
KEYWORDS   keratin; keratin 10; polymorphism; suprabasal keratin.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 45)
AUTHORS    Rothnagel, J.J., Dominey, A., Fisher, M., Axtell, S., Pittelkow, M.,
            Anton-Lamprecht, I., Hohl, D. and Roop, D.

```

**TITLE** Identification of mutational hot spots in the suprabasal keratin genes from patients with epidermolytic hyperkeratosis  
**JOURNAL** Unpublished (1993)  
**FEATURES** Location/Qualifiers  
 source  
 1..45  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="Lymphocyte"  
 /dev\_stage="Adult"  
 /tissue\_type="Blood"  
 /map="17q21-q23"  
 /note="from EHK patient"  
 1..45  
 /gene="KRT10"  
 <1..>45  
 /gene="KRT10"  
 /note="helix-initiation motif of 1A segment of rod domain"  
 /codon\_start=1  
 /function="intermediate filament precursor"  
 /db\_xref="GDB:G00-118-828"  
 /product="keratin 10"  
 /protein\_id="AAB59438.1"  
 /db\_xref="GI:307088"  
 /translation="KVTQMQLNDHLASYL"  
 29  
 /gene="KRT10"  
 /note="polymorphism results in an Arg->His mutation at amino acid position #10 in the HK10 rod; G00-118-828"  
 /phenotype="wild-type (g)"  
 /replace="g"  
 14 a 12 c 8 g 11 t  
 BASE COUNT  
 ORIGIN  
 Query Match 64.0%; Score 12.8; DB 97; Length 45;  
 Best Local Similarity 87.5%; Pred. No. 6.2e+04;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 5 aaccagatggtcattc 20  
 ||||| |||||  
 Db 36 ACCCAGGTGTCATTC 21  
 RESULT 5  
 LOCUS I64891 35 bp DNA PAT 07-OCT-1997  
 DEFINITION Sequence 7 from patent US 5665874.  
 ACCESSION I64891  
 VERSION I64891.1 GI:2481785  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 35)  
 AUTHORS Kuhajda,F.P. and Pasternack,G.R.  
 TITLE Cancer related antigen  
 JOURNAL Patent: US 5665874-A 7 09-SEP-1997;  
 FEATURES Location/Qualifiers  
 source  
 1..35  
 /organism="unknown"  
 9 a 12 c 6 g 8 t  
 BASE COUNT  
 ORIGIN  
 Query Match 63.0%; Score 12.6; DB 10; Length 35;  
 Best Local Similarity 78.9%; Pred. No. 7.9e+04;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 ttcaaccagatggtcattc 20  
 ||||| ||||| |||||  
 Db 35 TTCAAGAAGATGCCATGC 17

RESULT 6  
 LOCUS E13249/c 41 bp DNA PAT 24-JUN-1998  
 DEFINITION Oligonucleotide for microgene construction.  
 ACCESSION E13249  
 VERSION E13249.1 GI:3252054  
 KEYWORDS JP 1997154585-A/9.  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 41)  
 AUTHORS Shiba,K..  
 TITLE FORMATION OF RANDOM POLYMER OF MICROGENE  
 JOURNAL Patent: JP 1997154585-A 9 17-JUN-1997;  
 COMMENT KAGAKU GIJUTSU SHINKO JIGYODAN  
 OS None  
 OC Artificial sequences.  
 PN JP 1997154585-A/9  
 PD 17-JUN-1997  
 PF 06-DEC-1995 JP 1995318396  
 PI SHIBA KIYOTAKA  
 PC C12N15/09,C07H21/04//C12N9/00;  
 CC strandedness: Single;  
 CC topology: Linear;  
 FH Key  
 FT Location/Qualifiers  
 1..41  
 /organism="Artificial sequences".  
 FEATURES Location/Qualifiers  
 source  
 1..41  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 10 a 10 c 15 g 6 t  
 BASE COUNT  
 ORIGIN  
 Query Match 63.0%; Score 12.6; DB 10; Length 41;  
 Best Local Similarity 78.9%; Pred. No. 7.9e+04;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 tttaaccagatggtcatt 19  
 ||||| ||||| |||||  
 Db 32 TTTCACCATGATCCCTT 14  
 RESULT 7  
 LOCUS AR004965 30 bp DNA PAT 04-DEC-1998  
 DEFINITION Sequence 4 from patent US 5747312.  
 ACCESSION AR004965  
 VERSION AR004965.1 GI:3965844  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 30)  
 AUTHORS Wei,Y. and Sutton,G.G. III.  
 TITLE Human ALKB polypeptide  
 JOURNAL Patent: US 5747312-A 4 05-MAY-1998;  
 FEATURES Location/Qualifiers  
 source  
 1..30  
 /organism="unknown"  
 9 a 9 c 7 g 5 t  
 BASE COUNT  
 ORIGIN  
 Query Match 62.0%; Score 12.4; DB 9; Length 30;  
 Best Local Similarity 92.9%; Pred. No. 1e+05;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 tttaaccagatggtg 14  
 ||||| ||||| |||||  
 Db 9 TTTCATCCAGATGG 22

RESULT 8  
 LOCUS I40158 30 bp DNA PAT 13-MAY-1997  
 DEFINITION Sequence 4 from patent US 5618717.  
 ACCESSION I40158  
 VERSION I40158.1 GI:2083163  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 30)  
 AUTHORS Wei,Y. and Sutton,G.G. III.  
 TITLE DNA encoding human AlkB  
 JOURNAL Patent: US 5618717-A 4 08-APR-1997;  
 FEATURES Location/Qualifiers  
 source 1..30  
 BASE COUNT 9 a 9 c 7 g 5 t  
 ORIGIN  
  
 Query Match 62.0%; Score 12.4; DB 10; Length 30;  
 Best Local Similarity 92.9%; Pred. No. 1e+05;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 tttaaccagatgg 14  
 ||||| |||||  
 Db 9 TTTCATCCAGATGG 22  
  
 RESULT 9  
 LOCUS AR016116 20 bp DNA PAT 05-DEC-1998  
 DEFINITION Sequence 4 from patent US 5776682.  
 ACCESSION AR016116  
 VERSION AR016116.1 GI:3972393  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS First,M.Kent, Agoulnik,A.I. and Muallem,A.  
 TITLE Male infertility y-deletion detection battery  
 JOURNAL Patent: US 5776682-A 4 07-JUL-1998;  
 FEATURES Location/Qualifiers  
 source 1..20  
 BASE COUNT 4 a 6 c 3 g 7 t  
 ORIGIN  
  
 Query Match 61.0%; Score 12.2; DB 9; Length 20;  
 Best Local Similarity 82.4%; Pred. No. 1.3e+05;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 2 ttcaaccagatggtcat 18  
 || ||||| |||||  
 Db 1 TCCATCCAGCTGGTCAT 17  
  
 RESULT 10  
 LOCUS AR019114 20 bp DNA PAT 05-DEC-1998  
 DEFINITION Sequence 4 from patent US 5783390.  
 ACCESSION AR019114  
 VERSION AR019114.1 GI:3974228  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Svendsen,A., Patkar,S.Anant, Gormsen,E., Okkels,J.Sigurd and Thellersen,M.  
 TITLE Lipase variants  
 JOURNAL Patent: US 5869438-A 38 09-FEB-1999;  
 FEATURES Location/Qualifiers  
 source 1..33  
 BASE COUNT 10 a 7 c 3 g 13 t  
 ORIGIN

AUTHORS First,M.Kent and Agoulnik,A.I.  
 TITLE Male infertility y-deletion detection battery  
 JOURNAL Patent: US 5783390-A 4 21-JUL-1998;  
 FEATURES Location/Qualifiers  
 source 1..20  
 BASE COUNT 4 a 6 c 3 g 7 t  
 ORIGIN  
  
 Query Match 61.0%; Score 12.2; DB 9; Length 20;  
 Best Local Similarity 82.4%; Pred. No. 1.3e+05;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 2 ttcaaccagatggtcat 18  
 || ||||| |||||  
 Db 1 TCCATCCAGCTGGTCAT 17  
  
 RESULT 11  
 LOCUS AR090906 28 bp DNA PAT 07-SEP-2000  
 DEFINITION Sequence 1026 from patent US 5994076.  
 ACCESSION AR090906  
 VERSION AR090906.1 GI:10017661  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 28)  
 AUTHORS Chenchik,A., Johhadze,G. and Bibilashvili,R.  
 TITLE Methods of assaying differential expression  
 JOURNAL Patent: US 5994076-A 1026 30-NOV-1999;  
 FEATURES Location/Qualifiers  
 source 1..28  
 BASE COUNT 6 a 8 c 5 g 9 t  
 ORIGIN  
  
 Query Match 61.0%; Score 12.2; DB 9; Length 28;  
 Best Local Similarity 82.4%; Pred. No. 1.3e+05;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 2 ttcaaccagatggtcat 18  
 || ||||| |||||  
 Db 3 TTCTACCAGGTGTCAT 19  
  
 RESULT 12  
 LOCUS AR034456 33 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 38 from patent US 5869438.  
 ACCESSION AR034456  
 VERSION AR034456.1 GI:5950061  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 33)  
 AUTHORS Svendsen,A., Patkar,S.Anant, Gormsen,E., Okkels,J.Sigurd and Thellersen,M.  
 TITLE Lipase variants  
 JOURNAL Patent: US 5869438-A 38 09-FEB-1999;  
 FEATURES Location/Qualifiers  
 source 1..33  
 BASE COUNT 10 a 7 c 3 g 13 t  
 ORIGIN  
  
 Query Match 61.0%; Score 12.2; DB 9; Length 33;  
 Best Local Similarity 82.4%; Pred. No. 1.3e+05;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttccaaccagatggtca 17  
|||||  
Db 10 TTTCAACCAAGAGTAA 26

## RESULT 13

AR070401 LOCUS AR070401 33 bp DNA PAT 18-FEB-2000

DEFINITION Sequence 38 from patent US 5892013.

ACCESSION AR070401

VERSION AR070401.1 GI:7221289

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 33)

AUTHORS Svendsen,A., Patkar,S.,Anant, Gormsen,E., Clausen,I.,Groth,

TITLE Okkels,J.,Sigurd and Thellersen,M.

JOURNAL Lipase variants

PATENT: US 5892013-A 38 06-APR-1999;

FEATURES Location/Qualifiers

source 1..33

BASE COUNT 10 a 7 c 3 g 13 t

ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 33;

Best Local Similarity 82.4%; Pred. No. 1.3e+05;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttccaaccagatggtca 17

|||||

Db 10 TTTCAACCAAGAGTAA 26

## RESULT 14

AR026654/c

LOCUS AR026654 34 bp DNA PAT 29-SEP-1999

DEFINITION Sequence 9 from patent US 5856126.

ACCESSION AR026654

VERSION AR026654.1 GI:5937494

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 34)

AUTHORS Fukuchi,N., Yamamoto,H., Nagano,M., Kito,M., Tanaka,A., Ishii,K.,

TITLE Kobayashi,T. and Yoshimoto,R.

Peptide having anti-thrombus activity and method of producing the

same

JOURNAL Patent: US 5856126-A 9 05-JAN-1999;

FEATURES Location/Qualifiers

source 1..34

BASE COUNT 5 a 7 c 12 g 10 t

ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 34;

Best Local Similarity 82.4%; Pred. No. 1.3e+05;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 caaccagatggtcattc 20

|||||

Db 34 CAACCGGAGGACATTC 18

## RESULT 15

AX067989/c

LOCUS AX067989 29 bp DNA PAT 19-JAN-2001

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:07 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-18

Perfect score: 20

Sequence: 1 tttaaccagatggtcattc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est33.\*
- 26: gb\_est34.\*
- 27: gb\_est35.\*
- 28: gb\_est36.\*
- 29: gb\_est37.\*
- 30: gb\_est38.\*
- 31: gb\_est39.\*
- 32: gb\_est40.\*
- 33: em\_estba.\*
- 34: em\_estfun.\*
- 35: em\_esthum1.\*
- 36: em\_esthum2.\*
- 37: em\_esthum3.\*
- 38: em\_esthum4.\*
- 39: em\_esthum5.\*
- 40: em\_esthum6.\*
- 41: em\_esthum7.\*
- 42: em\_esthum8.\*
- 43: em\_esthum9.\*

- 44: em\_esthum10.\*
- 45: em\_esthum11.\*
- 46: em\_esthum12.\*
- 47: em\_esthum13.\*
- 48: em\_esthum14.\*
- 49: em\_esthum15.\*
- 50: em\_esthum16.\*
- 51: em\_esthum17.\*
- 52: em\_esthum18.\*
- 53: em\_esthum19.\*
- 54: em\_esthum20.\*
- 55: em\_esthum21.\*
- 56: em\_esthum22.\*
- 57: em\_esthum23.\*
- 58: em\_esthum24.\*
- 59: em\_esthum25.\*
- 60: em\_esthum26.\*
- 61: em\_esthum27.\*
- 62: em\_esthum28.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estin5.\*
- 68: em\_estom1.\*
- 69: em\_estom2.\*
- 70: em\_estov1.\*
- 71: em\_estov2.\*
- 72: em\_estpl1.\*
- 73: em\_estpl2.\*
- 74: em\_estpl3.\*
- 75: em\_estpl4.\*
- 76: em\_estpl5.\*
- 77: em\_estpl6.\*
- 78: em\_estpl7.\*
- 79: em\_estpl8.\*
- 80: em\_estpl9.\*
- 81: em\_estpl10.\*
- 82: em\_estro1.\*
- 83: em\_estro2.\*
- 84: em\_estro3.\*
- 85: em\_estro4.\*
- 86: em\_estro5.\*
- 87: em\_estro6.\*
- 88: em\_estro7.\*
- 89: em\_estro8.\*
- 90: em\_estro9.\*
- 91: em\_estro10.\*
- 92: em\_estro11.\*
- 93: em\_estro12.\*
- 94: em\_estro13.\*
- 95: em\_estro14.\*
- 96: em\_estro15.\*
- 97: em\_estro16.\*
- 98: em\_estro17.\*
- 99: em\_estro18.\*
- 100: em\_estro19.\*
- 101: em\_estro20.\*
- 102: gb\_est25.\*
- 103: gb\_est26.\*
- 104: gb\_est27.\*
- 105: gb\_est28.\*
- 106: gb\_est29.\*
- 107: gb\_est30.\*
- 108: gb\_est31.\*
- 109: gb\_est32.\*
- 110: gb\_est41.\*
- 111: gb\_est42.\*
- 112: gb\_est43.\*
- 113: gb\_est44.\*
- 114: gb\_est45.\*
- 115: gb\_est46.\*
- 116: gb\_est47.\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
Insert Length: 1015 Std Error: 0.00  
Seq primer: 40UP from Gibco

High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source

1. .25  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1836621"  
/clone.lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc. and primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAATGGGAGCGCGCCCAATTTTTTTTTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 5 a 8 c 4 g 8 t  
ORIGIN

Query Match 62.0%; Score 12.4; DB 17; Length 25;  
Best Local Similarity 92.9%; Pred. No. 4.4e+04;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 caaccagatggtca 17  
||| |||||

Db 20 CAATCAGATGTC 7

RESULT 2

AZ429959/c 39 bp DNA GSS 03-OCT-2000  
LOCUS  
DEFINITION  
clone U0214N08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0214N08 F, DNA sequence.

ACCESSION  
AZ429959

VERSION  
GSS.

KEYWORDS  
SOURCE

ORGANISM  
house mouse

REFERENCE  
AUTHORS

TITLE  
Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah  
Genome Center  
Rm. 308, Biomedical  
Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	12.4	62.0	25	17	AI208968	q929f11.x
c 2	12.4	62.0	39	243	AZ429959	IM0214N08
c 3	12.2	61.0	50	107	AU104028	AU104028
c 4	12.2	61.0	50	107	AU104139	AU104139
c 5	12	60.0	43	242	AZ388563	AZ388563
c 6	12	60.0	49	16	AI125432	q93d12.x
c 7	12	60.0	49	241	AZ311307	IM0026K11
c 8	11.6	58.0	31	247	AZ453622	IM0255E08
c 9	11.6	58.0	34	247	AZ663664	IM0543J20
c 10	11.6	58.0	35	247	AZ631220	IM0485J13
c 11	11.6	58.0	36	4	AA266819	mz91a02.r
c 12	11.6	58.0	43	2	AA075907	zm74e08.r
c 13	11.6	58.0	46	249	AZ777948	2M0012P13
c 14	11.2	56.0	34	16	AI096045	SMOVL3CAN
c 15	11.2	56.0	42	13	AA936922	om55c08.s
c 16	11.2	56.0	42	249	AZ799790	2M0057M22
c 17	11.2	56.0	43	244	AZ475536	IM0293P23
c 18	11.2	56.0	44	11	AA725427	aj13e09.s
c 19	11.2	56.0	49	10	AA700915	zj44c07.s
c 20	11.2	56.0	50	107	AU104136	AU104136
c 21	11.2	56.0	50	107	AU104138	AU104138
c 22	11	55.0	24	241	AZ317925	IM0036G20
c 23	11	55.0	25	244	AZ449662	IM0248D06
c 24	11	55.0	28	243	AZ427748	IM0209M22
c 25	11	55.0	35	154	BG504713	602551919
c 26	11	55.0	35	250	AZ821491	2M0094H18
c 27	11	55.0	45	189	T80630	yd92c01.s1
c 28	11	55.0	49	2	AA075329	zm86b10.s
c 29	11	55.0	49	107	AA053904	AU053904
c 30	11	55.0	50	31	AV535125	AV535125
c 31	10.8	54.0	29	243	AZ396455	IM0161F06
c 32	10.8	54.0	34	244	AZ451127	IM0250C16
c 33	10.8	54.0	40	14	AA990517	ua63906.s
c 34	10.8	54.0	42	242	AZ388234	IM0148E12
c 35	10.8	54.0	43	243	AZ428834	IM0212F07
c 36	10.8	54.0	44	158	H55075	CHR220014 C
c 37	10.8	54.0	50	114	AW285566	LGI_245_D
c 38	10.6	53.0	26	244	AZ463718	IM0272D16
c 39	10.6	53.0	26	249	AZ780163	2M0017D18
c 40	10.6	53.0	26	249	AZ783434	2M0025G12
c 41	10.6	53.0	30	241	AZ320504	IM0040P03
c 42	10.6	53.0	34	11	AA734076	vs19b04.r
c 43	10.6	53.0	34	13	AA929647	vy81f07.r
c 44	10.6	53.0	34	249	AZ792823	2M0045L06
c 45	10.6	53.0	35	258	TA262E04P	AL484164 T. brucei

## ALIGNMENTS

RESULT	1	25 bp	EST	29-NOV-1998
LOCUS	AI208968/c	q929f11.x1	Soares_testis_NHT	Hom sapiens cDNA clone IMAGE:1836621
DEFINITION		3' similar to SW:TEA-LYTP1 P02553	TUBULIN ALPHA CHAIN ; mRNA	
ACCESSION	AI208968			
VERSION	AI208968.1	GI:3770910		
KEYWORDS	EST			
SOURCE	human			
ORGANISM	Homo sapiens			
REFERENCE				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS				1 (bases 1 to 25)
TITLE				NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0214 row: N column: 08  
 Seq primer: CGTTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 39.

# FEATURES

Location/Qualifiers  
 1..39  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0214N08"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: pWD42nv: Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

4 a 8 c 15 g 12 t

# BASE COUNT

Query Match 62.0%; Score 12.4; DB 243; Length 39;  
 Best Local Similarity 92.9%; Pred. No. 4.7e+04;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tttaaccagatgg 14  
 ||| |||||

Db 20 TTGGAACCATGG 7

# RESULT

AU104028 50 bp mRNA EST 05-APR-2001  
 AU104028 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 ADSE00222, mRNA sequence.  
 ACCESSION AU104028  
 VERSION AU104028.1 GI:13553549  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 50)  
 AUTHORS Suzuki.Y., Tsunoda.T., Taira.H., Mizushima-Sugano.J., Sese.J., Hata  
 ,K., Suyama.A. and Sugano.S.  
 H., Ota.T., Isogai.T., Tanaka.T., Nakamura.Y., Morishita,S., Okubo  
 S. Construction and characterization of a full length-enriched and

TITLE Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL Unpublished (2001)  
 COMMENT Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: ysuzuki@ims.u-tokyo.ac.jp  
 Suzuki.Y., Yoshitomo-Nakagawa.K., Maruyama.K., Suyama.A. and Sugano  
 S. Construction and characterization of a full length-enriched and

a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  
 FEATURES Location/Qualifiers  
 1..50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="ADSE00222"  
 /clone\_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 8 a 10 c 16 g 16 t  
 ORIGIN

Query Match 61.0%; Score 12.2; DB 107; Length 50;  
 Best Local Similarity 82.4%; Pred. No. 6.2e+04;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tttaaccagatgg 17  
 ||| |||||

Db 13 TTTCATCCTGTGGTCA 29

# RESULT

AU104139 50 bp mRNA EST 05-APR-2001  
 AU104139 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 HEP21348, mRNA sequence.  
 ACCESSION AU104139  
 VERSION AU104139.1 GI:13553660  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 50)  
 AUTHORS Suzuki.Y., Tsunoda.T., Taira.H., Mizushima-Sugano.J., Sese.J., Hata  
 ,K., Ota.T., Isogai.T., Tanaka.T., Nakamura.Y., Morishita,S., Okubo  
 H., Suyama.A. and Sugano.S.  
 Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL Unpublished (2001)  
 COMMENT Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: ysuzuki@ims.u-tokyo.ac.jp  
 Suzuki.Y., Yoshitomo-Nakagawa.K., Maruyama.K., Suyama.A. and Sugano  
 S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES Location/Qualifiers  
 1..50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HEP21348"  
 /clone\_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 10 a 10 c 15 g 15 t  
 ORIGIN

# Query Match

Best Local Similarity 82.4%; Score 12.2; DB 107; Length 50;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tttaaccagatgg 18  
 ||| |||||

Db 9 TTCAACCCAGGAGGTGAT 25

# RESULT

AZ388563/c 43 bp DNA GSS 02-OCT-2000  
 LOCUS AZ388563  
 DEFINITION IM0148F08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0148F08 R, DNA sequence.

ACCESSION AZ388563  
 VERSION AZ388563.1 GI:10502271

ACCESSION VERSION KEYWORDS SOURCE ORGANISM	REFERENCE AUTHORS TITLE	JOURNAL COMMENT	FEATURES SOURCE
--	-------------------------------	--------------------	--------------------

REGIONAL SOURCE	BASE COUNT ORIGIN
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

DB	RESULT	7	20
	AZ311307		
	LOCUS		
	DEFINITION		
	ACCESSION		
	VERSION		
	KEYWORDS		
	SOURCE		
	ORGANISM		

## REFERENCE AUTHORS

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0026 row: K column: 11  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 49.

## FEATURES

source

1. .49  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0026K11"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (gil14732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

## BASE COUNT

8 a 10 c 8 g 23 t

## ORIGIN

Query Match 60.0%; Score 12; DB 241; Length 49;  
 Best Local Similarity 75.0%; Pred. No. 7.8e+04;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttccaaccagatggcttc 20

Db 12 TTTCACCCAGATTCTATTC 31

## RESULT

8

LOCUS A2453622 31 bp DNA GSS 04-OCT-2000  
 DEFINITION IM0255E08F Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0255E08 F, DNA sequence.

ACCESSION A2453622

VERSION A2453622.1 GI:10611604

KEYWORDS GSS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 31)

## AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0255 row: E column: 08  
 Seq primer: CGTCTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 31.

## FEATURES

source

1. .31  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0255E08"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (gil14732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

## BASE COUNT

9 a 8 c 7 g 7 t

## ORIGIN

Query Match 58.0%; Score 11.6; DB 244; Length 31;  
 Best Local Similarity 77.8%; Pred. No. 1.2e+05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttccaaccagatggctcat 18

Db 18 TTTCACCCAGATGGGCAT 1

## RESULT

9

LOCUS A2663664/c 34 bp DNA GSS 14-DEC-2000  
 DEFINITION IM0543J20F Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0543J20 F, DNA sequence.

ACCESSION A2663664

VERSION A2663664.1 GI:11800810

KEYWORDS GSS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 34)

## AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0543 row: J column: 20  
 Seq primer: CGTTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 34.

# FEATURES

1. .34  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0543J20"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 8 a 7 c 7 g 12 t  
 ORIGIN

Query Match 58.0%; Score 11.6; DB 247; Length 34;  
 Best Local Similarity 77.8%; Pred. No. 1.2e+05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 tcaaccagatggtcattc 20  
 | ||||| ||||| |||||  
 Db 23 TGAACCAATGGCGATTC 6

# RESULT 10

AZ631220  
 LOCUS AZ631220 35 bp DNA GSS 13-DEC-2000  
 DEFINITION clone UUGC1M0485J13 F, DNA sequence.  
 ACCESSION AZ631220  
 VERSION AZ631220.1 GI:11753410  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 35)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

# TITLE JOURNAL COMMENT

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0485 row: J column: 13  
 Seq primer: CGTTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 35.

# FEATURES

1. .35  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0485J13"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 14 a 6 c 5 g 10 t  
 ORIGIN

Query Match 58.0%; Score 11.6; DB 247; Length 35;  
 Best Local Similarity 77.8%; Pred. No. 1.2e+05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tttaaccagatggtcat 18  
 | ||||| ||||| |||||  
 Db 15 TTTCAGGAGATGATCCT 32

# RESULT 11

AA266819/c  
 LOCUS AA266819 36 bp mRNA EST 21-MAR-1997  
 DEFINITION mz91a02.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone IMAGE:720746 5' similar to gb:D10061 Mouse mRNA for DNA topoisomerase I (MOUSE);, mRNA sequence.  
 ACCESSION AA266819  
 VERSION AA266819.1 GI:1903574  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
1 (bases 1 to 36)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Underwood,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996)

TITLE  
JOURNAL  
COMMENT  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:446242  
Trace considered overall poor quality  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 1.

FEATURES  
source

Location/Qualifiers  
1..36  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:720746"  
/clone\_lib="Soares mouse lymph node NbMLN"  
/sex="male"  
/tissue\_type="lymph node"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"

/note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGGAGCGCGCGATCTTTTTTTTTTTTTTTTTTTTT  
3'] ; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT  
ORIGIN  
18 a 4 c 8 g 6 t  
Query Match 58.0%; Score 11.6; DB 4; Length 36;  
Best Local Similarity 77.8%; Pred. No. 1.2e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttcaaccagatggtcatt 19  
|| |||| |||| ||||  
Db 27 TTGGACCAATGCTCATT 10

RESULT 12  
LOCUS  
AA075907/c 43 bp mRNA EST 23-DEC-1997  
DEFINITION  
zm74e08.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA  
clone IMAGE:531398 5' similar to TR:G972436 G972436 CYCLIN  
DEPENDENT PROTEIN KINASE, TYPE 4 ; mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
AA075907  
AA075907.1 GI:1615777  
EST.  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
AUTHORS  
1 (bases 1 to 43)  
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissole,S., Dietrich,N., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Wardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

WARNING: There is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 1.

FEATURES  
source

Location/Qualifiers  
1..43  
/organism="Homo sapiens"  
/db\_xref="GDB:3920910"  
/db\_xref="taxon:9606"  
/clone="IMAGE:531398"  
/clone\_lib="Stratagene neuroepithelium (#937231)"  
/dev\_stage="Ntera-2/RA neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="vector: pBluescript SK; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2 cells (Ntera-2/ci.D1) induced with Retinoic Acid for 24 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCCAGCAG 3' -3' adaptor sequence: 5' CTCGACGTTTTTTTTTTTTTTT 3"

BASE COUNT  
ORIGIN  
8 a 8 c 14 g 13 t

Query Match 58.0%; Score 11.6; DB 2; Length 43;  
Best Local Similarity 77.8%; Pred. No. 1.2e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttcaaccagatggtcat 18  
||||||| |||||  
Db 32 TTTCACCACTGGCTCAT 15

RESULT 13  
LOCUS  
AA777948/c 46 bp DNA GSS 16-FEB-2001  
DEFINITION  
2M0012P13R Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGC2M0012P13 R, DNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
AA777948  
AA777948.1 GI:12907079  
GSS.  
house mouse.

ORGANISM  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 46)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.

REFERENCE  
AUTHORS  
1 (bases 1 to 46)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT 84112, USA

The library was constructed by Wenhong Lu. The library is available from Dr. S.A. Williams, email [genome@smith.edu](mailto:genome@smith.edu) When requesting this clone from Dr. Williams, please reference the Williams lab clone id - SMOVL3CAN20F05

Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 1.

```

1..34
Location/Qualifiers
/organism="Onchocerca volvulus"
/strain="Sierra Leone"
/db_xref="taxon:6282"
/clone="onch1306"
/clone_lib="Onchocerca volvulus infective larva cDNA
(SAW94WL-OVL3)"
/lab_host="Y1-blue Mdp'"

```

/lab\_host="XLI-Blue MRF"  
 /name="Vector: lambda Unizap XR; Site\_1: EcoR I; Site\_2:  
 Xho I; Cutaneous filarial nematode parasite of humans.  
 mRNA was prepared from third stage infective larvae of  
*Onchocerca volvulus* isolated from mosquitoes 10 days after  
 infection and converted to double stranded cDNA using  
 reverse transcriptase and oligo(dT) followed by RNase H  
 and DNAPol I. The library had 1.8 x 10E5 independent  
 recombinants and average insert size was 900 base pairs.  
 The library was constructed by Wenhong Lu. The library is  
 available from Dr. S.A. Williams, email genome@smith.edu."

Query Match 56.0%; Score 11.2; DB 16; Length 34;  
Best Local Similarity 81.2%; Pred. No. 1.9e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels

Qy	4	caaccagatggtcatt	19
Db	27	CAATCAAAATGCTCGTT	12

RESULT 15  
AA936922

[illegible]

SOURCE	ORGANISM
Italian.	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 42)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

**JOURNAL COMMENT**  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
Insert Length: 753 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.

FEATURES  
source

Location/Qualifiers  
1. .42  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1551086"  
/clone\_lib="NCI\_CGAP\_GC4"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from 3 pooled  
germ cell tumors, and was then primed with a Not I -  
oligo(47) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT7T3  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 14 a 7 c 11 g 10 t  
ORIGIN

Query Match 56.0%; Score 11.2; DB 13; Length 42;  
Best Local Similarity 81.2%; Pred. No. 1.9e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 3 tcaaccagatggctcat 18  
||| | |||||  
Db 22 TCATCATGATGTCAT 37

Search completed: October 2, 2001, 15:01:08  
Job time: 10837 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 12:00:36 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-3

Perfect score: 20

Sequence: 1 ccgcgggtcacagtgtcg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*\*

1: gb\_bal:\*\*

2: gb\_ba2:\*\*

3: gb\_ba3:\*\*

4: gb\_in1:\*\*

5: gb\_in2:\*\*

6: gb\_in3:\*\*

7: gb\_om:\*\*

8: gb\_ov:\*\*

9: gb\_pat1:\*\*

10: gb\_pat2:\*\*

11: gb\_ph:\*\*

12: gb\_p11:\*\*

13: gb\_p12:\*\*

14: gb\_p13:\*\*

15: gb\_p14:\*\*

16: em\_bal:\*\*

17: em\_ba2:\*\*

18: em\_fun:\*\*

19: em\_htgo\_hum:\*\*

20: em\_htgo\_inv:\*\*

21: em\_htgo\_rod:\*\*

22: em\_htg\_hum1:\*\*

23: em\_htg\_hum2:\*\*

24: em\_htg\_hum3:\*\*

25: em\_htg\_hum4:\*\*

26: em\_htg\_hum5:\*\*

27: em\_htg\_hum6:\*\*

28: em\_htg\_hum7:\*\*

29: em\_htg\_hum8:\*\*

30: em\_htg\_inv1:\*\*

31: em\_htg\_inv2:\*\*

32: em\_htg\_other:\*\*

33: em\_htg\_rod:\*\*

34: em\_hum1:\*\*

35: em\_hum2:\*\*

36: em\_hum3:\*\*

37: em\_hum4:\*\*

38: em\_hum5:\*\*

39: em\_hum6:\*\*

40: em\_hum7:\*\*

41: em\_in:\*\*

42: em\_om:\*\*

43: em\_or:\*\*

44: em\_ov:\*\*  
45: em\_pat:\*\*  
46: em\_ph:\*\*  
47: em\_pl:\*\*  
48: em\_ro:\*\*  
49: em\_sts:\*\*  
50: em\_sy:\*\*  
51: em\_un:\*\*  
52: em\_vi:\*\*  
53: gb\_sts1:\*\*  
54: gb\_sts2:\*\*  
55: gb\_sts3:\*\*  
56: gb\_sy:\*\*  
57: gb\_un:\*\*  
58: gb\_vil:\*\*  
59: gb\_v12:\*\*  
60: gb\_htg1:\*\*  
61: gb\_htg2:\*\*  
62: gb\_htg3:\*\*  
63: gb\_htg4:\*\*  
64: gb\_htg5:\*\*  
65: gb\_htg6:\*\*  
66: gb\_htg7:\*\*  
67: gb\_htg8:\*\*  
68: gb\_htg9:\*\*  
69: gb\_htg10:\*\*  
70: gb\_htg11:\*\*  
71: gb\_htg12:\*\*  
72: gb\_htg13:\*\*  
73: gb\_htg14:\*\*  
74: gb\_htg15:\*\*  
75: gb\_htg16:\*\*  
76: gb\_htg17:\*\*  
77: gb\_htg18:\*\*  
78: gb\_htg19:\*\*  
79: gb\_htg20:\*\*  
80: gb\_htg21:\*\*  
81: gb\_htg22:\*\*  
82: gb\_htg23:\*\*  
83: gb\_htg24:\*\*  
84: gb\_htg25:\*\*  
85: gb\_pr1:\*\*  
86: gb\_pr2:\*\*  
87: gb\_pr3:\*\*  
88: gb\_pr4:\*\*  
89: gb\_pr5:\*\*  
90: gb\_pr6:\*\*  
91: gb\_pr7:\*\*  
92: gb\_pr8:\*\*  
93: gb\_pr9:\*\*  
94: gb\_rol:\*\*  
95: gb\_ro2:\*\*  
96: gb\_in4:\*\*  
97: gb\_pr10:\*\*  
98: em\_ba3:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	13.2	66.0	33	10	I92389
2	13.2	66.0	50	9	AX043850
3	12.6	63.0	37	10	AX090386
4	12.6	63.0	41	10	AX090365
5	12.4	62.0	45	94	MWTCRAIC7
6	12.2	61.0	29	9	AX026028
7	12.2	61.0	30	10	E28465
8	12.2	61.0	40	9	AR095613
					I92389 Sequence 41
					AX043850 Sequence
					AX090386 Sequence
					AX090365 Sequence
					X70713 M.musculus
					AX026028 Sequence
					E28465 Hyaluronate
					AR095613 Sequence

9 12.2 61.0 40 9 AR095615  
 10 12.2 61.0 41 9 AX043844  
 11 12.2 61.0 44 10 AX090361  
 12 12.2 61.0 45 9 AX043846  
 13 11.8 59.0 16 9 AR041138  
 14 11.8 59.0 22 10 I43055  
 15 11.8 59.0 29 9 AR041128  
 16 11.8 59.0 30 9 AR041129  
 17 11.8 59.0 30 9 AR041130  
 18 11.8 59.0 35 10 I31649  
 19 11.8 59.0 36 10 I21472  
 20 11.8 59.0 44 10 I43051  
 21 11.6 58.0 24 9 AR017555  
 22 11.6 58.0 24 10 I21733  
 23 11.6 58.0 25 9 A58772  
 24 11.6 58.0 25 10 I19529  
 25 11.6 58.0 25 10 I19530  
 26 11.6 58.0 25 10 I19532  
 27 11.6 58.0 25 10 I86209  
 28 11.6 58.0 25 10 I86210  
 29 11.6 58.0 25 10 I86212  
 30 11.6 58.0 26 9 A58770  
 31 11.6 58.0 26 9 AR017553  
 32 11.6 58.0 26 10 I19534  
 33 11.6 58.0 26 10 I21731  
 34 11.6 58.0 26 10 I86214  
 35 11.6 58.0 30 9 AR049374  
 36 11.6 58.0 30 9 AR095535  
 37 11.6 58.0 30 9 AX033429  
 38 11.6 58.0 50 9 A58773  
 39 11.4 57.0 18 9 AR078873  
 40 11.4 57.0 22 10 I76917  
 41 11.4 57.0 25 9 A65881  
 42 11.4 57.0 25 9 AX042765  
 43 11.4 57.0 37 9 S45430  
 44 11.2 56.0 20 9 AR067323  
 45 11.2 56.0 20 9 AR073339

## ALIGNMENTS

RESULT 1  
 192389  
 LOCUS I92389 33 bp DNA PAT 01-DEC-1998  
 DEFINITION Sequence 41 from patent US 5728519.  
 ACCESSION I92389  
 VERSION I92389.1 GI:3936859  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 33)  
 AUTHORS Levenbook,I.S., Chumakov,K.M., Norwood,L.P. and Roninson,I.  
 TITLE Assay for virulent revertants of attenuated live vaccines and kits therefor  
 JOURNAL Patent: US 5728519-A 41 17-MAR-1998;  
 FEATURES Location/Qualifiers  
 source 1..33  
 BASE COUNT 6 a 7 c 9 g 11 t  
 ORIGIN  
 Query Match 66.0%; Score 13.2; DB 10; Length 33;  
 Best Local Similarity 83.3%; Pred. No. 5.2e+04;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcgggtcacagtgg 18  
 ||||| ||| ||| ||  
 Db 9 CCGCGGGTTCAAAGTAGT 26

RESULT 2  
 AX043850  
 LOCUS AX043850 50 bp DNA PAT 23-NOV-2000  
 DEFINITION Sequence 57 from Patent WO0063391.  
 ACCESSION AX043850  
 VERSION AX043850.1 GI:11342435  
 KEYWORDS  
 ORGANISM synthetic construct.  
 SOURCE synthetic construct.  
 REFERENCE 1 (bases 1 to 50)  
 AUTHORS Savidge,B., Lassner,M.W., Weiss,J.D. and Post-Beitemmiller,D.  
 TITLE Nucleic acid sequences to proteins involved in tocopherol synthesis  
 JOURNAL Patent: WO 0063391-A 57 26-OCT-2000;  
 FEATURES Location/Qualifiers  
 source 1..50  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="Oligonucleotide"

BASE COUNT 9 a 16 c 14 g 11 t  
 ORIGIN  
 Query Match 66.0%; Score 13.2; DB 9; Length 50;  
 Best Local Similarity 83.3%; Pred. No. 4.7e+04;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcgggtcacagtgg 18  
 ||||| ||| ||| |||  
 Db 5 CCGCGGCCGCAATGGT 22

RESULT 3  
 AX090386  
 LOCUS AX090386 37 bp DNA PAT 21-MAR-2001  
 DEFINITION Sequence 79 from Patent WO0116308.  
 ACCESSION AX090386  
 VERSION AX090386.1 GI:13444245  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM synthetic construct.  
 REFERENCE 1 (bases 1 to 37)  
 AUTHORS Lassner,M. and van Penennaam,A.  
 TITLE Plant sterol acyltransferases  
 JOURNAL Patent: WO 0116308-A 79 08-MAR-2001;  
 FEATURES Location/Qualifiers  
 source 1..37  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="Synthetic oligonucleotide primer"

BASE COUNT 8 a 12 c 11 g 6 t  
 ORIGIN  
 Query Match 63.0%; Score 12.6; DB 10; Length 37;  
 Best Local Similarity 78.9%; Pred. No. 1e+05;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ccgcgggtcacagtgg 19  
 ||||| ||| ||| |||  
 Db 5 CCGCGGCCGCAATGGC 23

RESULT 4  
 AX090365  
 LOCUS AX090365 41 bp DNA PAT 21-MAR-2001  
 DEFINITION Sequence 58 from Patent WO0116308.  
 ACCESSION AX090365  
 VERSION AX090365.1 GI:13444226  
 KEYWORDS

SOURCE  
ORGANISM synthetic construct.  
artificial sequence.

REFERENCE  
1 (bases 1 to 41)  
Lassner, M. and van Eenennaam, A.  
TITLE Plant sterol acyltransferases  
JOURNAL Patent: WO 0116308-A 58 08-MAR-2001;  
MONSANTO COMPANY (US)  
FEATURES Location/Qualifiers  
source 1..41  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Synthetic oligonucleotide primer"  
BASE COUNT 5 a 13 c 15 g 8 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 41;  
Best Local Similarity 78.9%; Pred. No. 9.9e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccgcgggctcacagtgtc 19  
||||| | |||| |||| |  
Db 5 CCgcggccgcacacatggc 23

RESULT 5  
MMTCRAIC7/c  
LOCUS M.musculus mRNA for T-cell receptor, V-J alpha junction (A3/IIC7). 15-MAR-1993  
DEFINITION X70713  
ACCESSION X70713  
VERSION X70713.1 GI:288017  
KEYWORDS J-alpha gene segment; junction; T-cell receptor; V-alpha gene segment.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 45)  
AUTHORS Casanova, J.  
TITLE Direct Submission  
JOURNAL Submitted (15-JAN-1993) J. Casanova, INSERM U132, Developpement Normal et Pathologique du Systeme Immunitaire, Necker, Pavillon Kirmisson, 149 rue de Sevres, 75743 Paris Cedex, FRANCE  
REFERENCE 2 (bases 1 to 45)  
AUTHORS Casanova, J.L., Martinon, F., Gournier, H., Barra, C., Pannetier, C., Regnault, A., Kourilsky, P., Cerottini, J.C. and Maryanski, J.L.  
TITLE T cell receptor selection by and recognition of two class I major histocompatibility complex-restricted antigenic peptides that differ at a single position  
JOURNAL J. Exp. Med. 177 (3), 811-820 (1993)  
MEDLINE 93171821  
FEATURES Location/Qualifiers  
source 1..45  
/organism="Mus musculus"  
/strain="DBA2"  
/isolate="A3/IIC7"  
/db\_xref="taxon:10090"  
/haplotype="H2-d"  
/tissue\_type="T cell"  
/cell\_type="CTL"  
<1..>45  
mRNA  
BASE COUNT 6 a 9 c 18 g 12 t  
ORIGIN

Query Match 62.0%; Score 12.4; DB 94; Length 45;  
Best Local Similarity 92.9%; Pred. No. 1.2e+05;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgcgggctcacag 14  
||||| |||||||

Db 18 CCgcggcgtcacag 5

RESULT 6  
AX026028/c  
LOCUS AX026028 29 bp DNA PAT 16-SEP-2000  
DEFINITION Sequence 9 from Patent WO0031239.  
ACCESSION AX026028  
VERSION AX026028.1 GI:10187483  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 29)  
AUTHORS Eshhar, Z., Bolhuis, R.L. and Willemse, R.A.  
TITLE Immune cells having predefined biological specificity, comprising chimeric T cell receptor  
JOURNAL Patent: WO 0031239-A 9 02-JUN-2000;  
ESHAR ZELIG (IL); YEDA RES & DEV (IL); BOLHUIS REINDER L H (NL); WILLEMSSEN RALPH A (NL)  
FEATURES Location/Qualifiers  
source 1..29  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Synthetic DNA"  
BASE COUNT 4 a 10 c 11 g 4 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 29;  
Best Local Similarity 82.4%; Pred. No. 1.7e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 cgggctcacagtgtcg 20  
||||| ||||| ||||| |||||  
Db 24 CCgcgtcacagtgtcg 8

RESULT 7  
E28465/c  
LOCUS E28465 30 bp DNA PAT 07-FEB-2001  
DEFINITION Hyaluronate synthase promoter DNA.  
ACCESSION E28465  
VERSION E28465.1 GI:13018357  
KEYWORDS JP 1999196875-A/13.  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Yoichi, Y., N.I.I. and Kimata.  
TITLE Hyaluronate synthase promoter DNA  
JOURNAL Patent: JP 1999196875-A 13 27-JUL-1999;  
SEIKAGAKU KOGYO CO LTD  
COMMENT OS Unidentified  
PN JP 1999196875-A/13  
PD 27-JUL-1999  
PF 14-JAN-1998 JP 1998006191  
PR YOICHI YAMADA, NAOKI ITANO, KOJI KIMATA  
PI C12N15/09, C12N9/00, C12Q1/68, C12N15/09, C12R1/91, C12N15/00,  
PC (C12N15/00, C12R1/91)  
CC Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..30  
/organism="Unidentified".  
FEATURES Location/Qualifiers  
source 1..30  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 7 a 11 c 7 g 5 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 10; Length 30;  
Best Local Similarity 82.4%; Pred. No. 1.7e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ccgcgggctcacagtgg 18  
||||| ||||| |||||  
DB 18 CCGGTGCTCATGTGT 2

RESULT 8  
AR095613  
LOCUS AR095613 40 bp DNA PAT 08-SEP-2000  
DEFINITION Sequence 8 from patent US 6004793.  
ACCESSION AR095613  
VERSION AR095613.1 GI:10023640  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 40)  
AUTHORS Lunnen,K.D., Dalton,M.A., Wilson,G.G. and Xu,S.  
TITLE Method for cloning and producing the AwaI restriction endonuclease in E. coli and purification of the recombinant AwaI restriction endonuclease  
JOURNAL Patent: US 6004793-A 8 21-DEC-1999;  
FEATURES Location/Qualifiers  
source 1..40  
/organism="unknown"  
BASE COUNT 12 a 12 c 9 g 7 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 40;  
Best Local Similarity 82.4%; Pred. No. 1.6e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccgcgggctcacagtgg 17  
||| ||||| |||||  
DB 8 CCGCGGCTCAAGTTG 24

RESULT 9  
AR095615  
LOCUS AR095615 40 bp DNA PAT 08-SEP-2000  
DEFINITION Sequence 10 from patent US 6004793.  
ACCESSION AR095615  
VERSION AR095615.1 GI:10023644  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 40)  
AUTHORS Lunnen,K.D., Dalton,M.A., Wilson,G.G. and Xu,S.  
TITLE Method for cloning and producing the AwaI restriction endonuclease in E. coli and purification of the recombinant AwaI restriction endonuclease  
JOURNAL Patent: US 6004793-A 10 21-DEC-1999;  
FEATURES Location/Qualifiers  
source 1..40  
/organism="unknown"  
BASE COUNT 12 a 12 c 9 g 7 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 40;  
Best Local Similarity 82.4%; Pred. No. 1.6e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccgcgggctcacagtgg 17  
||| ||||| |||||  
DB 8 CCGCGGCTCAAGTTG 24

RESULT 10  
AX043844  
LOCUS AX043844 41 bp DNA PAT 23-NOV-2000  
DEFINITION Sequence 51 from Patent WO0063391.  
ACCESSION AX043844  
VERSION AX043844.1 GI:11342429  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 41)  
AUTHORS Savidge,B., Lassner,M.W., Weiss,J.D. and Post-Beittemiller,D.  
TITLE Nucleic acid sequences to proteins involved in tocopherol synthesis  
JOURNAL Patent: WO 0063391-A 51 26-OCT-2000;  
FEATURES Location/Qualifiers  
source 1..41  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Oligonucleotide"  
BASE COUNT 6 a 13 c 11 g 11 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 41;  
Best Local Similarity 82.4%; Pred. No. 1.6e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccgcgggctcacagtgg 17  
||||| ||||| |||||  
DB 5 CCGCGGCGCACATGG 21

RESULT 11  
AX090361  
LOCUS AX090361 44 bp DNA PAT 21-MAR-2001  
DEFINITION Sequence 54 from Patent WO0116308.  
ACCESSION AX090361  
VERSION AX090361.1 GI:13444222  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 44)  
AUTHORS Lassner,M. and van Eenennaam,A.  
TITLE Plant sterol acyltransferases  
JOURNAL Patent: WO 0116308-A 54 08-MAR-2001;  
FEATURES Location/Qualifiers  
source 1..44  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Synthetic oligonucleotide primer"  
BASE COUNT 13 a 11 c 14 g 6 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 10; Length 44;  
Best Local Similarity 82.4%; Pred. No. 1.6e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccgcgggctcacagtgg 17  
||||| ||||| |||||  
DB 5 CCGCGGCGCACATGG 21

RESULT 12  
AX043846  
LOCUS AX043846 45 bp DNA PAT 23-NOV-2000  
DEFINITION Sequence 53 from Patent WO0063391.  
ACCESSION AX043846  
VERSION AX043846.1 GI:11342431

KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE artificial sequence.  
1 (bases 1 to 45)  
AUTHORS Savidge,B., Lassner,M.W., Weiss,J.D. and Post-Belittenmiller,D.  
TITLE Nucleic acid sequences to proteins involved in tocopherol synthesis  
JOURNAL Patent: WO.0063391-A 53 26-OCT-2000;  
Calgene LLC (US)

FEATURES  
source.  
1. .45  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Oligonucleotide"  
BASE COUNT 4 a 13 c 14 g 14 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 45;  
Best Local Similarity 82.4%; Pred. No. 1.5e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ccgcgggctcacagtgg 17  
||||| | ||||| |||  
Db 5 CCGCGCCGCAATGG 21

RESULT 13  
AR041138/c  
LOCUS AR041138 16 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 13 from patent US 5811269.  
ACCESSION AR041138  
VERSION AR041138.1 GI:5961634  
KEYWORDS  
SOURCE Unknown.  
ORGANISM  
REFERENCE 1 (bases 1 to 16)  
AUTHORS Nadeau,J.G., Dean,C.H., Schram,J.L., Howard,D.R., Dey,M.S. and Wright,D.J.  
TITLE Detection of mycobacteria by multiplex nucleic acid amplification  
JOURNAL Patent: US 5811269-A 13 22-SEP-1998;  
FEATURES Location/Qualifiers  
source  
1. .16  
/organism="unknown"  
BASE COUNT 3 a 3 c 6 g 4 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 16;  
Best Local Similarity 86.7%; Pred. No. 3.1e+05;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ccgcgggctcacagt 15  
||||| | ||||| |||  
Db 15 CCGCATGCTCACAGT 1

RESULT 14  
I43055  
LOCUS I43055 22 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 38 from patent US 5631130.  
ACCESSION I43055  
VERSION I43055.1 GI:2468299  
KEYWORDS  
SOURCE Unknown.  
ORGANISM  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Leckie,G.W., Davis,A.H., Semple-Facey,I.E., Manlove,M.T. and Solomon,N.A.  
TITLE Materials and methods for the detection of Mycobacterium tuberculosis

JOURNAL Patent: US 5631130-A 38 20-MAY-1997;  
FEATURES Location/Qualifiers  
source  
1. .22  
/organism="unknown"  
BASE COUNT 4 a 9 c 4 g 5 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 10; Length 22;  
Best Local Similarity 86.7%; Pred. No. 2.9e+05;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ccgcgggctcacagt 15  
||||| | ||||| |||  
Db 7 CCGCACGCTCACAGT 21

RESULT 15  
AR041128  
LOCUS AR041128 29 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 3 from patent US 5811269.  
ACCESSION AR041128  
VERSION AR041128.1 GI:5961624  
KEYWORDS  
SOURCE Unknown.  
ORGANISM  
REFERENCE 1 (bases 1 to 29)  
AUTHORS Nadeau,J.G., Dean,C.H., Schram,J.L., Howard,D.R., Dey,M.S. and Wright,D.J.  
TITLE Detection of mycobacteria by multiplex nucleic acid amplification  
JOURNAL Patent: US 5811269-A 3 22-SEP-1998;  
FEATURES Location/Qualifiers  
source  
1. .29  
/organism="unknown"  
BASE COUNT 7 a 9 c 7 g 6 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 29;  
Best Local Similarity 86.7%; Pred. No. 2.7e+05;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ccgcgggctcacagt 15  
||||| | ||||| |||  
Db 14 CCGCATGCTCACAGT 28

Search completed: October 2, 2001, 15:56:35  
Job time: 14159 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 12:00:31 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-3

Perfect score: 20

Sequence: 1 ccgcggggtcacagtgtcg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_est3: \*  
4: gb\_est4: \*  
5: gb\_est5: \*  
6: gb\_est6: \*  
7: gb\_est7: \*  
8: gb\_est8: \*  
9: gb\_est9: \*  
10: gb\_est10: \*  
11: gb\_est11: \*  
12: gb\_est12: \*  
13: gb\_est13: \*  
14: gb\_est14: \*  
15: gb\_est15: \*  
16: gb\_est16: \*  
17: gb\_est17: \*  
18: gb\_est18: \*  
19: gb\_est19: \*  
20: gb\_est20: \*  
21: gb\_est21: \*  
22: gb\_est22: \*  
23: gb\_est23: \*  
24: gb\_est24: \*  
25: gb\_est33: \*  
26: gb\_est34: \*  
27: gb\_est35: \*  
28: gb\_est36: \*  
29: gb\_est37: \*  
30: gb\_est38: \*  
31: gb\_est39: \*  
32: gb\_est40: \*  
33: em\_estba: \*  
34: em\_estfun: \*  
35: em\_esthum1: \*  
36: em\_esthum2: \*  
37: em\_esthum3: \*  
38: em\_esthum4: \*  
39: em\_esthum5: \*  
40: em\_esthum6: \*  
41: em\_esthum7: \*  
42: em\_esthum8: \*  
43: em\_esthum9: \*

44: em\_esthum10: \*  
45: em\_esthum11: \*  
46: em\_esthum12: \*  
47: em\_esthum13: \*  
48: em\_esthum14: \*  
49: em\_esthum15: \*  
50: em\_esthum16: \*  
51: em\_esthum17: \*  
52: em\_esthum18: \*  
53: em\_esthum19: \*  
54: em\_esthum20: \*  
55: em\_esthum21: \*  
56: em\_esthum22: \*  
57: em\_esthum23: \*  
58: em\_esthum24: \*  
59: em\_esthum25: \*  
60: em\_esthum26: \*  
61: em\_esthum27: \*  
62: em\_esthum28: \*  
63: em\_estin1: \*  
64: em\_estin2: \*  
65: em\_estin3: \*  
66: em\_estin4: \*  
67: em\_estin5: \*  
68: em\_estom1: \*  
69: em\_estom2: \*  
70: em\_estov1: \*  
71: em\_estov2: \*  
72: em\_estpl1: \*  
73: em\_estpl2: \*  
74: em\_estpl3: \*  
75: em\_estpl4: \*  
76: em\_estpl5: \*  
77: em\_estpl6: \*  
78: em\_estpl7: \*  
79: em\_estpl8: \*  
80: em\_estpl9: \*  
81: em\_estpl10: \*  
82: em\_estro1: \*  
83: em\_estro2: \*  
84: em\_estro3: \*  
85: em\_estro4: \*  
86: em\_estro5: \*  
87: em\_estro6: \*  
88: em\_estro7: \*  
89: em\_estro8: \*  
90: em\_estro9: \*  
91: em\_estro10: \*  
92: em\_estro11: \*  
93: em\_estro12: \*  
94: em\_estro13: \*  
95: em\_estro14: \*  
96: em\_estro15: \*  
97: em\_estro16: \*  
98: em\_estro17: \*  
99: em\_estro18: \*  
100: em\_estro19: \*  
101: em\_estro20: \*  
102: gb\_est25: \*  
103: gb\_est26: \*  
104: gb\_est27: \*  
105: gb\_est28: \*  
106: gb\_est29: \*  
107: gb\_est30: \*  
108: gb\_est31: \*  
109: gb\_est32: \*  
110: gb\_est41: \*  
111: gb\_est42: \*  
112: gb\_est43: \*  
113: gb\_est44: \*  
114: gb\_est45: \*  
115: gb\_est46: \*  
116: gb\_est47: \*





## SUMMARIES

Result No.	Score	Query %			DB	ID	Description
		Match	Length				
1	14.4	72.0	50	107	AU106328	AU106328 AU106328	
2	13.6	68.0	46	3	AA196905	AA196905 zq09B06.i	
3	13.2	66.0	50	107	AU107932	AU107932 AU107932	
4	12.8	64.0	40	244	A2453043	A2453043 1M0254A13	
5	12.8	64.0	41	12	AA822940	AA822940 vp30g03.r	
6	12.6	63.0	25	24	A1762402	A1762402 wh65e10.x	
7	12.6	63.0	31	12	AA8445818	AA8445818 ak85f01.s	
8	12.6	63.0	46	244	A2465895	A2465895 1M0276F12	
9	12.6	63.0	50	250	A28334614	A28334614 2M0117121	
10	12	60.0	44	258	TA233A12Q	TA233A12Q T. brucei	
11	12	60.0	50	107	AU106948	AU106948 AU106948	
12	11.6	58.0	22	258	TA140A04P	TA140A04P T. brucei	
13	11.6	58.0	28	24	A1735009	A1735009 as44e04.x	
14	11.6	58.0	37	7	AA466917	AA466917 vf10g04.r	
15	11.6	58.0	46	9	AA607432	AA607432 vm61h02.f	
16	11.6	58.0	46	107	AU106912	AU106912 AU106912	
17	11.6	58.0	50	16	AA149037	AA149037 qc75d09.x	
18	11.4	57.0	29	249	A2766277	A2766277 1M0563A15	
19	11.4	57.0	50	107	AU104386	AU104386 AU104386	
20	11.2	56.0	25	258	TA205H11Q	TA205H11Q T. brucei	
21	11.2	56.0	31	242	A2345566	A2345566 1M0080I14	
22	11.2	56.0	37	23	A1684291	A1684291 tw97b09.x	
23	11.2	56.0	37	250	A2812781	A2812781 2M0079B17	
24	11.2	56.0	38	175	BG292849	BG292849 602389872	
25	11.2	56.0	50	107	AU102352	AU102352 AU102352	
26	11.2	56.0	50	107	AU102356	AU102356 AU102356	
27	11.2	56.0	50	107	AU104505	AU104505 AU104505	
28	11	55.0	32	243	A2404438	A2404438 1M0173116	
29	11	55.0	41	247	A2647118	A2647118 1M0513023	
30	11	55.0	44	166	BE311212	BE311212 601057922	
31	11	55.0	46	144	BF123084	BF123084 601761846	
32	11	55.0	46	242	A2388569	A2388569 1M0148G09	
33	11	55.0	50	107	AU103426	AU103426 AU103426	
34	11	55.0	50	107	AU103606	AU103606 AU103606	
35	11	55.0	50	107	AU103608	AU103608 AU103608	
36	11	55.0	50	107	AU103612	AU103612 AU103612	
37	11	55.0	50	107	AU103614	AU103614 AU103614	
38	11	55.0	50	107	AU103617	AU103617 AU103617	
39	11	55.0	50	107	AU104542	AU104542 AU104542	
40	11	55.0	50	107	AU107007	AU107007 AU107007	
41	10.8	54.0	25	249	A2770438	A2770438 1M0572G03	
42	10.8	54.0	27	244	A2455866	A2455866 1M0258K21	
43	10.8	54.0	27	244	A2463607	A2463607 1M0272K02	
44	10.8	54.0	35	245	A2497161	A2497161 1M0233G22	
45	10.8	54.0	38	246	A2591206	A2591206 1M0401K03	

## ALIGNMENTS

RESULT	1	
AUI06328		
LOCUS	50 bp	EST
DEFINITION	Sugano Homo sapiens cDNA library Homo sapiens cDNA clone	
ACCESSION	AUI06328	KAT00883, mRNA sequence.
VERSION	AUI06328.1	GI:13555849
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 50)	
AUTHORS	Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata	
	,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo	
	,K., Suyama,A. and Sugano,S.	
TITLE	Fine structural analysis of transcription start sites of human	

Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally.  
 Primer: Oligo dt. Skeletal muscle from patient with  
 malignant hyperthermia. Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG  
 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT  
 ORIGIN

12 a 15 c 14 g 5 t

Query Match 68.0%; Score 13.6; DB 3; Length 46;  
 Best Local Similarity 80.0%; Pred. No. 1.2e+04;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccgagggtcacagtgtgc 20  
 |||| |||| |||| ||||  
 Db 41 CCGCCGCTCAATGTGTGTC 22

RESULT 3

AU107932 50 bp mRNA EST 05-APR-2001  
 LOCUS AU107932 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 DEFINITION ZR62008, mRNA sequence.

ACCESSION AU107932

VERSION AU107932.1 GI:13557454

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 50)

AUTHORS Suzuki.Y., Tsunoda.T., Taira.H., Mizushima-Sugano.J., Sese,J., Hata

.K., Suyama.A. and Sugano.S.

TITLE Fine Structural analysis of transcription start sites of human

mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL Department of Virology

COMMENT Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: yszuk@iims.u-tokyo.ac.jp

Suzuki.Y., Yoshitomo-Nakagawa.K., Maruyama.K., Suyama.A. and Sugano

.S. Construction and characterization of a full length-enriched and

a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES Location/Qualifiers

1..50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="ZR62008"

/clone\_lib="Sugano Homo sapiens cDNA library"

6 a 19 c 15 g 9 t 1 others

BASE COUNT 6 a 19 c 15 g 9 t 1 others

ORIGIN

Query Match 66.0%; Score 13.2; DB 107; Length 50;

Best Local Similarity 83.3%; Pred. No. 1.9e+04;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ccgagggtcacagtgtgc 19

||||| ||||| ||||| |||||

Db 17 CTCGGCGGCCAGTGTGTC 34

RESULT 4

AZ453043/c 40 bp DNA GSS 04-OCT-2000

LOCUS IM0254A13F Mouse 10kb plasmid UUGCLM library Mus musculus genomic

DEFINITION clone UUGCLM0254A13 F, DNA sequence.

ACCESSION AZ453043

VERSION AZ453043.1 GI:10610442

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0254 row: A column: 13

Seg primer: CGTTCTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 40.

Location/Qualifiers

1..40

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGCLM0254A13"

/clone\_lib="Mouse 10kb plasmid UUGCLM library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 12 a 11 c 9 g 8 t

ORIGIN

Query Match 64.0%; Score 12.8; DB 244; Length 40;

Best Local Similarity 87.5%; Pred. No. 3e+04;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ccgagggtcacagtgc 17

||||| ||||| ||||| |||||

Db 26 CGGGGCGCTCACAGTGG 11

RESULT 5

AA822940/c 41 bp mRNA EST 17-FEB-1998

LOCUS vp30g03.r1 Barstead mouse proximal colon MPLR86 Mus musculus cDNA

DEFINITION clone IMAGE:1078228 5' similar to gb:U13705 Mus musculus domesticus

C57BL/6J plasma glutathione (MOUSE);, mRNA sequence.

ACCESSION AA822940

VERSION AA822940.1 GI:2892808



IMAGE Consortium (info@image.llnl.gov) for further information.  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: ~40ml3 fwd. Et from Amersham  
 High quality sequence stop: 1.

## FEATURES

source

```

1. .31
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1414681"
/clone_lib="Barstead spleen HPLRB2"
/sex="male"
/dev_stage="adult, 17 years"
/lab_host="DH10B"
/note="Organ: spleen; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACGATCTGAAGTGGAGCGGCCCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGGATCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."

```

## BASE COUNT

ORIGIN

6 a 9 c 10 g 6 t

Query Match 63.0%; Score 12.6; DB 12; Length 31;  
 Best Local Similarity 78.9%; Pred. No. 3.7e+04; Mismatches 4; Indels 0; Gaps 0;  
 Matches 15; Conservative 0;

QY 1 ccgcgggtcacagtggtc 19

```

1 ||||| ||| |||||
Db 23 CAGCGCGTCAAAAGTGTC 5

```

## RESULT 8

AZ465895/c

LOCUS

IM0276F12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0276F12 F, DNA sequence.

ACCESSION

AZ465895

VERSION

GSS.

KEYWORDS

SOURCE

ORGANISM

house mouse.  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 46)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112 USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0276 row: F column: 12  
 Seq primer: CGTTGTAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 46.

## FEATURES

source

```

1. .46
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```

```

/clone="UUGC1M0276F12"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 [g114732114[gblAF129072.1], a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

```

BASE COUNT

ORIGIN

11 a 14 c 14 g 7 t

Query Match 63.0%; Score 12.6; DB 244; Length 46;

Best Local Similarity 78.9%; Pred. No. 3.8e+04;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccgcgggtcacagtggtc 19

```

1 ||||| ||| |||||
Db 33 CCTCGAGCTCACAGATC 15

```

## RESULT 9

AZ834614/c

LOCUS

2M0117121F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC2M0117121 F, DNA sequence.

ACCESSION

AZ834614

VERSION

GSS.

KEYWORDS

SOURCE

house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 50)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0117 row: I column: 21  
 Seq primer: CGTTGTAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 50.

## FEATURES

source

```

1. .50
/organism="Mus musculus"
/strain="C57BL/6J"

```

```

/db_xref="taxon:10090"
/clonelib="UUCG2M0117121"
/clonelib="Mouse 10kb plasmid UUCG1M library"
/lab_host="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydromatically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g114732114/gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      14 a      10 c      23 g      3 t
ORIGIN

Query Match      63.0%; Score 12.6; DB 250; Length 50;
Best Local Similarity 78.9%; Pred. No. 3.8e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtggctc 19
   | ||||| ||||| |||||
Db 30 CGCGGGCTCACTCGGAC 12

RESULT 10
TA233A12Q
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 233a12, reverse sequence, genomic survey sequence.
ACCESSION
AL481198
VERSION
AL481198.1 GI:11846892
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei.
ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE
1 (bases 1 to 44)
AUTHORS
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE
Direct Submission
JOURNAL
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhlesanger.ac.uk
COMMENT
Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTAT 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nhlesanger@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
Location/Qualifiers
source
1..44
/organism="Trypanosoma brucei"

```

```

/db_xref="TREU927"
/db_xref="taxon:5691"
/clonelib="233a12"
BASE COUNT      12 a      10 c      12 g      10 t
ORIGIN

Query Match      60.0%; Score 12; DB 258; Length 44;
Best Local Similarity 75.0%; Pred. No. 7.5e+04;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtggctc 20
   | ||||| ||||| |||||
Db 14 CAGCGGATCACAGGTACG 33

RESULT 11
AU106948/c
LOCUS
DEFINITION
AU106948 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
CAS03735, mRNA sequence.
ACCESSION
AU106948
VERSION
AU106948.1 GI:13556469
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 50)
AUTHORS
Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo, K., Suyama, A. and Sugano, S.
TITLE
Fine structural analysis of transcription start sites of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL
Unpublished (2001)
COMMENT
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yzuzuki@ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES
Location/Qualifiers
source
1..50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clonelib="CAS03735"
/clonelib="Sugano Homo sapiens cDNA library"
BASE COUNT      6 a      16 c      24 g      4 t
ORIGIN

Query Match      60.0%; Score 12; DB 107; Length 50;
Best Local Similarity 75.0%; Pred. No. 7.5e+04;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtggctc 20
   | ||||| ||||| |||||
Db 36 CCGCGGGCGCGCAGCAGCTCG 17

RESULT 12
TA140A04P/c
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 140a04, forward sequence, genomic survey sequence.
ACCESSION
AL466405
VERSION
AL466405.1 GI:11835760
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei.
ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

```



/dev\_stage="embryo (pre-implantation)"  
/lab\_host="DH10B"  
/notes="Organ: embryo; Vector: pSPORT; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally from mRNA prepared  
from 800 blastocysts. Primer: SalI(df): cDNAs were  
5'-CGGTGACCGTCGACCGTCTTTT-3'. cDNAs were  
cloned into the NotI/SalI sites of a pSPORT vector (Life  
Technologies). Two different size selections: B1 (larger  
inserts) and B3."

BASE COUNT 10 a 9 c 12 g 6 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 7; Length 37;  
Best Local Similarity 77.8%; Pred. No. 1.2e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtgt 18  
||| ||||| |||  
Db 37 CCGTGGGCTCACTCAGGT 20

RESULT 15  
AA607492/c  
LOCUS  
DEFINITION  
AA607492 46 bp mRNA EST 30-SEP-1997  
vm6lh02.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone  
IMAGE:1002771 5' similar to SW:RS9\_RAT P29314 40S RIBOSOMAL PROTEIN  
S9... mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AA607492.1 GI:2456385  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 46)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMNI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:566987

Possible reversed clone: similarity on wrong strand  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 1.

FEATURES  
SOURCE  
1..46  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1002771"  
/clone\_lib="Stratagene mouse Tcell 937311"  
/tissue\_type="Tcell"  
/dev\_stage="M30 CD4+ cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: blood; Vector: pBluescript SK-; Site 1:  
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. M30 CD4+ cells. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'."

BASE COUNT 10 a 13 c 14 g  
ORIGIN

Query Match 58.0%; Score 11.6; DB 9; Length 46;  
Best Local Similarity 77.8%; Pred. No. 1.2e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 ccgcgggctcacagtgt 18  
||| ||| ||||| |||  
Db 46 CCGTGGGCTCACATAGGT 29

Search completed: October 2, 2001, 15:00:45  
Job time: 10814 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:35 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-4  
Perfect score: 20  
Sequence: 1 ggcgcgtgaagcgaaggca 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_ba1:\*
  - 2: gb\_ba2:\*
  - 3: gb\_ba3:\*
  - 4: gb\_in1:\*
  - 5: gb\_in2:\*
  - 6: gb\_in3:\*
  - 7: gb\_om:\*
  - 8: gb\_ov:\*
  - 9: gb\_pat1:\*
  - 10: gb\_pat2:\*
  - 11: gb\_ph:\*
  - 12: gb\_pl1:\*
  - 13: gb\_pl2:\*
  - 14: gb\_pl3:\*
  - 15: gb\_pl4:\*
  - 16: em\_ba1:\*
  - 17: em\_ba2:\*
  - 18: em\_fun:\*
  - 19: em\_htgo\_hum:\*
  - 20: em\_htgo\_inv:\*
  - 21: em\_htgo\_rod:\*
  - 22: em\_htg\_hum1:\*
  - 23: em\_htg\_hum2:\*
  - 24: em\_htg\_hum3:\*
  - 25: em\_htg\_hum4:\*
  - 26: em\_htg\_hum5:\*
  - 27: em\_htg\_hum6:\*
  - 28: em\_htg\_hum7:\*
  - 29: em\_htg\_hum8:\*
  - 30: em\_htg\_inv1:\*
  - 31: em\_htg\_inv2:\*
  - 32: em\_htg\_other:\*
  - 33: em\_htg\_rod:\*
  - 34: em\_hum1:\*
  - 35: em\_hum2:\*
  - 36: em\_hum3:\*
  - 37: em\_hum4:\*
  - 38: em\_hum5:\*
  - 39: em\_hum6:\*
  - 40: em\_hum7:\*
  - 41: em\_in:\*
  - 42: em\_om:\*
  - 43: em\_or:\*

- 44: em\_ov:\*
- 45: em\_pat:\*
- 46: em\_ph:\*
- 47: em\_pl:\*
- 48: em\_ro:\*
- 49: em\_sts:\*
- 50: em\_sy:\*
- 51: em\_un:\*
- 52: em\_vi:\*
- 53: gb\_sts1:\*
- 54: gb\_sts2:\*
- 55: gb\_sts3:\*
- 56: gb\_sy:\*
- 57: gb\_un:\*
- 58: gb\_vil:\*
- 59: gb\_vil2:\*
- 60: gb\_htg1:\*
- 61: gb\_htg2:\*
- 62: gb\_htg3:\*
- 63: gb\_htg4:\*
- 64: gb\_htg5:\*
- 65: gb\_htg6:\*
- 66: gb\_htg7:\*
- 67: gb\_htg8:\*
- 68: gb\_htg9:\*
- 69: gb\_htg10:\*
- 70: gb\_htg11:\*
- 71: gb\_htg12:\*
- 72: gb\_htg13:\*
- 73: gb\_htg14:\*
- 74: gb\_htg15:\*
- 75: gb\_htg16:\*
- 76: gb\_htg17:\*
- 77: gb\_htg18:\*
- 78: gb\_htg19:\*
- 79: gb\_htg20:\*
- 80: gb\_htg21:\*
- 81: gb\_htg22:\*
- 82: gb\_htg23:\*
- 83: gb\_htg24:\*
- 84: gb\_htg25:\*
- 85: gb\_pr1:\*
- 86: gb\_pr2:\*
- 87: gb\_pr3:\*
- 88: gb\_pr4:\*
- 89: gb\_pr5:\*
- 90: gb\_pr6:\*
- 91: gb\_pr7:\*
- 92: gb\_pr8:\*
- 93: gb\_pr9:\*
- 94: gb\_rol:\*
- 95: gb\_ro2:\*
- 96: gb\_in4:\*
- 97: gb\_pr10:\*
- 98: em\_ba3:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	13.2	66.0	21	10	I27712
2	13.2	66.0	21	10	I83056
3	12.4	62.0	21	10	AX094895
c	4	12.2	61.0	22	A09940
	5	12.2	61.0	38	A47740
	6	12.2	61.0	38	A019411
c	7	12.2	61.0	47	HUMTCVD1DK
	8	12.2	61.0	50	AR092516
					I27712 Sequence 19
					I83056 Sequence 19
					AX094895 Sequence
					A09940 oligonucleo
					A47740 Sequence 4
					AR019411 Sequence
					L32460 Human (clon
					AR092516 Sequence



ORGANISM unidentified  
unclassified.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Moureau.P., Derclaye.I., Delor.I. and Cornelis.G.  
TITLE Nucleic acid probes useful for detecting specifically different bacterial species of the genus Campylobacter  
JOURNAL Patent: EP 0350392-A 2 10-JAN-1990;  
IRE-MEDGENIX S.A  
FEATURES Location/Qualifiers  
source 1..22  
/organism="unidentified"  
/db\_xref="taxon:32644" 7 t  
BASE COUNT 2 a 7 c 6 g  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 22;  
Best Local Similarity 82.4%; Pred. No. 2.7e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gccgcgtgaagcgaagg 18  
||||| ||||| |||||  
Db 17 GCTCGTGACCAAGG 1

RESULT 5  
LOCUS A47740 38 bp DNA PAT 07-MAR-1997  
DEFINITION Sequence 4 from Patent W09533833.  
ACCESSION A47740  
VERSION A47740.1 GI:2301650  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified.  
REFERENCE 1 (bases 1 to 38)  
AUTHORS Wood.P.C. and Quirk.A.V.  
TITLE YEAST STRAINS  
JOURNAL Patent: WO 9533833-A 4 14-DEC-1995;  
COMMENT DELTA BIOTECHNOLOGY LTD (GB)  
Other publication AU 2626295 960104.  
FEATURES Location/Qualifiers  
source 1..38  
/organism="unidentified"  
/db\_xref="taxon:32644" 9 t  
BASE COUNT 7 a 9 c 13 g  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 38;  
Best Local Similarity 82.4%; Pred. No. 2.4e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gccgtgaagcgaagca 20  
||||| ||||| |||||  
Db 1 GCCGTGAGCGAGGAA 17

RESULT 6  
LOCUS AR019411 38 bp DNA PAT 05-DEC-1998  
DEFINITION Sequence 4 from patent US 5783423.  
ACCESSION AR019411  
VERSION AR019411.1 GI:3974525  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
unclassified.  
REFERENCE 1 (bases 1 to 38)  
AUTHORS Wood.P.Carolyn and Quirk.A.Victor.  
TITLE Yeast strains  
JOURNAL Patent: US 5783423-A 4 21-JUL-1998;  
FEATURES Location/Qualifiers

source 1..38  
/organism="unknown"  
BASE COUNT 7 a 9 c 13 g  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 38;  
Best Local Similarity 82.4%; Pred. No. 2.4e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gccgtgaagcgaagca 20  
||||| ||||| |||||  
Db 1 GCCGTGAGCGAGGAA 17

RESULT 7  
LOCUS HUMTCVD1DK/c 47 bp mRNA PRI 10-FEB-1995  
DEFINITION Human (clone: 3cpj20) T-cell receptor delta-chain (V-delta-1) mRNA.  
ACCESSION L32460  
VERSION L32460.1 GI:497530  
KEYWORDS T-cell receptor; delta chain.  
SOURCE Homo sapiens intestine cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 47)  
AUTHORS Chowers.Y., Holtmeier.W., Harwood.J., Morzycka-Wroblewska,E. and Kagnoff.M.F.  
TITLE The V delta 1 T cell receptor repertoire in human small intestine and colon  
JOURNAL J. Exp. Med. 180 (1), 183-190 (1994)  
MEDLINE 94275371  
FEATURES Location/Qualifiers  
source 1..47  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="intestine" 14 t  
BASE COUNT 9 a 15 c 9 g  
ORIGIN

Query Match 61.0%; Score 12.2; DB 97; Length 47;  
Best Local Similarity 82.4%; Pred. No. 2.3e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gccgtgaagcgaagca 20  
||||| ||||| |||||  
Db 28 GCCATGTAGCGTAGGCA 12

RESULT 8  
LOCUS AR092516 50 bp DNA PAT 08-SEP-2000  
DEFINITION Sequence 46 from patent US 5998168.  
ACCESSION AR092516  
VERSION AR092516.1 GI:10019270  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
unclassified.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Loomore.S.M., Yacoub.R.Khayyam, Zealey,G.Ross and Klein,M.Henri.  
TITLE Expression of gene products from genetically manipulated strains of bordetella  
JOURNAL Patent: US 5998168-A 46 07-DEC-1999;  
FEATURES Location/Qualifiers  
source 1..50  
/organism="unknown" 4 t  
BASE COUNT 10 a 18 c 18 g  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 50;  
 Best Local Similarity 82.4%; Pred. No. 2.3e+05;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaag 17  
 |||||  
 Db 23 GCGCGGTGACGCGAAG 39

RESULT 9  
 I21584/c  
 LOCUS I21584 20 bp DNA PAT 07-OCT-1996  
 DEFINITION Sequence 4 from patent US 5521301.  
 ACCESSION I21584  
 VERSION I21584.1 GI:1601938  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Wallace,R,Bruce and Ugozzoli,L.  
 TITLE Genotyping of multiple allele systems  
 JOURNAL Patent: US 5521301-A 4 28-MAY-1996;  
 FEATURES Location/Qualifiers  
 1..20  
 /organism="unknown"  
 BASE COUNT 2 a 7 c 7 g 4 t  
 ORIGIN

Query Match 60.0%; Score 12; DB 10; Length 20;  
 Best Local Similarity 75.0%; Pred. No. 3.4e+05;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaagca 20  
 |||||  
 Db 20 GCGGCTACAGCGGTGCA 1

RESULT 10  
 A60848  
 LOCUS A60848 48 bp DNA PAT 06-MAR-1998  
 DEFINITION Sequence 157 from Patent WO9708320.  
 ACCESSION A60848  
 VERSION A60848.1 GI:3715469  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 48)  
 AUTHORS Knappik,A., Pack,P., Ilag,V., Ge,L., Moroney,S. and Plueckthun,A.  
 TITLE PROTEIN/(POLY)PEPTIDE LIBRARIES  
 JOURNAL Patent: WO 9708320-A 157 06-MAR-1997;  
 MORPHOSYS PROTEINOPTIMIERUNG (DE)  
 FEATURES Location/Qualifiers  
 1..48  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 9 a 14 c 17 g 8 t  
 ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 48;  
 Best Local Similarity 75.0%; Pred. No. 2.9e+05;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaagca 20  
 |||||  
 Db 1 GCGGCTAGAGCCCAAGCA 20

RESULT 11  
 AX020035

LOCUS AX020035 20 bp DNA PAT 07-SEP-2000  
 DEFINITION Sequence 49 from Patent WO9937764.  
 ACCESSION AX020035  
 VERSION AX020035.1 GI:10043864  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Veugelers,M.P. and David,G.J.  
 TITLE New members of the glypican gene family  
 JOURNAL Patent: WO 9937764-A 49 29-JUL-1999;  
 VEUGELERS MARK PAUL DITTMAR (BE); VLAAMS INTERUNIV INST BIOTECH  
 (BE); DAVID GUIDO JOSEPH FRANS (BE)  
 FEATURES Location/Qualifiers  
 1..20  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 9 a 3 c 6 g 2 t  
 ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 20;  
 Best Local Similarity 86.7%; Pred. No. 4.2e+05;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 cgtgaagcgaagcca 20  
 |||||  
 Db 5 CGTGAAGCAAGACA 19

RESULT 12  
 AX076946  
 LOCUS AX076946 25 bp DNA PAT 22-FEB-2001  
 DEFINITION Sequence 58 from Patent WO0105836.  
 ACCESSION AX076946  
 VERSION AX076946.1 GI:13121599  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM synthetic construct.  
 REFERENCE 1 (bases 1 to 25)  
 AUTHORS Botstein,D., Goddard,A., Gurney,A.L., Hillan,K.J., Roy,M.A. and  
 Wood,W.I.  
 TITLE Polypeptidic compositions and methods for the treatment of tumors  
 JOURNAL Patent: WO 0105836-A 58 25-JAN-2001;  
 Genentech, Inc. (US)  
 FEATURES Location/Qualifiers  
 1..25  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="Synthetic Oligonucleotide Probe"  
 BASE COUNT 9 a 8 c 5 g 3 t  
 ORIGIN

Query Match 58.0%; Score 11.6; DB 10; Length 25;  
 Best Local Similarity 77.8%; Pred. No. 5.1e+05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ggcgcgtgaagcgaagc 19  
 |||||  
 Db 3 GCGCGCAAGCAAGAC 20

RESULT 13  
 AR044492  
 LOCUS AR044492 27 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 32 from patent US 5817495.  
 ACCESSION AR044492  
 VERSION AR044492.1 GI:5965957  
 KEYWORDS

```

SOURCE      Unknown.
ORGANISM     Unclassified.
REFERENCE    1 (bases 1 to 27)
AUTHORS      Pedersen,A.Hjelholt, Vind,J., Svendsen,A., Cherry,J.R., Lamsa,M.,
              Schneider,P. and Jensen,B.Rostgaard.
TITLE        H.sub.2 O.sub.2 -stable peroxidase variants
JOURNAL      Patent: US 5817495-A 32 06-OCT-1998;
FEATURES     Location/Qualifiers
              1..27
              /organism="unknown"
BASE COUNT   5 a      9 c      11 g      2 t
ORIGIN

Query Match      58.0%; Score 11.6; DB 9; Length 27;
Best Local Similarity 77.8%; Pred. No. 5e+05;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggcgcctgaagcgaagg 18
    |||||  || |||||
Db 10 GGC CGCGCGAGAGAGG 27

RESULT 14
LOCUS      A93959          36 bp      DNA          PAT      22-JAN-2000
DEFINITION Sequence 13 from Patent WO9718322.
ACCESSION  A93959
VERSION     A93959.1 GI:6742060
KEYWORDS   .
SOURCE     unidentified.
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 36)
AUTHORS    Dahm,M.W.
TITLE      METHOD OF QUANTIFYING TUMOUR CELLS IN A BODY FLUID AND A SUITABLE
JOURNAL    TEST KIT
JOURNAL    Patent: WO 9718322-A 13 22-MAY-1997;
           DAHM MICHAEL W (DE)
FEATURES   Location/Qualifiers
              1..36
              /organism="unidentified"
              /db_xref="taxon:32644"
BASE COUNT 7 a      15 c      8 g      6 t
ORIGIN

Query Match      58.0%; Score 11.6; DB 9; Length 36;
Best Local Similarity 77.8%; Pred. No. 4.7e+05;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggcgcctgaagcgaagg 18
    |||||  || |||||
Db 18 GGCACGGGAATCGATCG 1

RESULT 15
LOCUS      AR012355          43 bp      DNA          PAT      04-DEC-1998
DEFINITION Sequence 46 from patent US 5763284.
ACCESSION  AR012355
VERSION     AR012355.1 GI:3970345
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 43)
AUTHORS    Tal,R., Wong,H.C., Casipit,C., Chavallaz,P. and Wittman,V.
TITLE      Methods for peptide synthesis and purification
JOURNAL    Patent: US 5763284-A 46 09-JUN-1998;
FEATURES   Location/Qualifiers
              1..43
              source

```

```

BASE COUNT      8 a      11 c      15 g      9 t
ORIGIN

Query Match      58.0%; Score 11.6; DB 9; Length 43;
Best Local Similarity 77.8%; Pred. No. 4.6e+05;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggcgcctgaagcgaagg 18
    |||||  || |||||
Db 1 GGGCCCATGGATCGAGG 18

Search completed: October 2, 2001, 15:56:37
Job time: 14161 sec

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:00:45 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-4

Perfect score: 20

Sequence: 1 ggcgcgtgaagcggaaggca 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_estl1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: gb\_est41.\*  
42: gb\_est42.\*  
43: gb\_est43.\*  
44: gb\_est44.\*  
45: gb\_est45.\*  
46: gb\_est46.\*  
47: gb\_est47.\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	14	70.0	50	107	AU103598
C 2	13.6	68.0	50	107	AU105195
C 3	13.6	68.0	50	107	AU105195
C 4	13.6	68.0	50	107	AU105197
C 5	13.2	66.0	47	250	AU105198
C 6	13.2	66.0	50	107	AU105198
C 7	12.8	64.0	50	107	AZ816650
C 8	12.8	64.0	50	107	2M085D03
C 9	12.8	64.0	50	107	AU105437
C 10	12.8	64.0	50	107	AU105437
C 11	12.8	64.0	50	107	AU103768
C 12	12.8	64.0	50	107	AU103768
C 13	12.8	64.0	50	107	AU103771
C 14	12.8	64.0	50	107	AU103771
C 15	12.8	64.0	50	107	AU103787
C 16	12.8	64.0	50	107	AU103787
C 17	12.8	64.0	50	107	AU103791
C 18	12.6	63.0	50	107	AU103974
C 19	12.4	62.0	50	107	AU103974
C 20	12.2	61.0	50	107	AU103975
C 21	12.2	61.0	50	107	AU103975
C 22	12.2	61.0	50	107	AU103977
C 23	12.2	61.0	50	107	AU103977
C 24	12.2	61.0	50	107	AU103977
C 25	12.2	61.0	50	107	AU103977
C 26	12.2	61.0	50	107	AU103977
C 27	12.2	61.0	50	107	AU103977
C 28	12.2	61.0	50	107	AU103977
C 29	12.2	61.0	50	107	AU103977
C 30	12.2	61.0	50	107	AU103977
C 31	11.8	59.0	46	14	AU103981
C 32	11.8	59.0	50	107	AU103981
C 33	11.8	59.0	50	107	AU103981
C 34	11.8	59.0	50	107	AU103981
C 35	11.8	59.0	50	107	AU103981
C 36	11.8	59.0	50	107	AU103981
C 37	11.8	59.0	50	107	AU103981
C 38	11.8	59.0	50	107	AU103981
C 39	11.8	59.0	50	107	AU103981
C 40	11.8	59.0	50	107	AU103981
C 41	11.8	59.0	50	107	AU103981
C 42	11.8	59.0	50	107	AU103981
C 43	11.8	59.0	50	107	AU103981
C 44	11.8	59.0	50	107	AU103981
C 45	11.8	59.0	50	107	AU103981

## ALIGNMENTS

RESULT 1  
 AU103598/c 50 bp mRNA EST 05-APR-2001  
 LOCUS AU103598 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 DEFINITION COLF6124, mRNA sequence.  
 ACCESSION AU103598  
 VERSION AU103598.1 GI:13553119  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 50)  
 AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.  
 TITLE Fine Structural analysis of transcription start sites of human

## JOURNAL COMMENT

MRNAs using full-length enriched and 5'-end enriched cDNA libraries  
 Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: yzuzuki@iems.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

source  
 1..50

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="COLF6124"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
 BASE COUNT 8 a 16 c 19 g 7 t  
 ORIGIN

Query Match 70.0%; Score 14; DB 107; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 6e+03;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ggcgcgtgaagcga 15  
 |||||

Db 18 GCGCGTGAAGCGA 5  
 |||||

## RESULT 2

AU105195 50 bp mRNA EST 05-APR-2001  
 LOCUS AU105195 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 DEFINITION HRC06645, mRNA sequence.  
 ACCESSION AU105195  
 VERSION AU105195.1 GI:13554716  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 50)  
 AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.

## TITLE

Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

## JOURNAL COMMENT

Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: yzuzuki@iems.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

source  
 1..50

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HRC06645"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
 BASE COUNT 8 a 13 c 24 g 5 t  
 ORIGIN

Query Match 68.0%; Score 13.6; DB 107; Length 50;  
 Best Local Similarity 80.0%; Pred. No. 9.5e+03;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcga 20  
 |||||

Db 1 GGAGCTGTGAGCGCAGGCA 20  
 |||||



10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 8 a 7 c 22 g 10 t  
ORIGIN

Query Match 66.0%; Score 13.2; DB 250; Length 47;  
Best Local Similarity 83.3%; Pred. No. 1.5e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgcgcgtgaagcgaagcga 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 38 CCCCGTGAAGAGAAAGCA 21

RESULT 6  
AUI05437/c

LOCUS AUI05437 50 bp mRNA EST 05-APR-2001  
DEFINITION AUI05437 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
HEP15178, mRNA sequence.  
ACCESSION AUI05437  
VERSION AUI05437.1 GI:13554958  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 50)  
Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata ,K., Suyama,A. and Sugano,S.  
Fine Structural analysis of transcription start sites of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL

COMMENT

Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yszukie@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano ,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1. .50  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP15178"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
2 a 21 c 15 g 12 t

Query Match 66.0%; Score 13.2; DB 107; Length 50;  
Best Local Similarity 83.3%; Pred. No. 1.5e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggcgcctgaagcgaagcga 18  
| | | | | | | | | | | | | | | | | | | | | |  
Db 35 GGCCCCCGAAGCGGAAG 18

RESULT 7  
AUI03765/c

LOCUS AUI03765 50 bp mRNA EST 05-APR-2001  
DEFINITION AUI03765 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
HEP11908, mRNA sequence.

ACCESSION AUI03765  
VERSION AUI03765.1 GI:13553286  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 50)  
Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata ,K., Suyama,A. and Sugano,S.  
Fine Structural analysis of transcription start sites of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL

COMMENT

Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yszukie@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano ,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1. .50  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP11908"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
11 a 14 c 12 g 13 t

Query Match 64.0%; Score 12.8; DB 107; Length 50;  
Best Local Similarity 87.5%; Pred. No. 2.4e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggcgcctgaagcgaagcga 16  
| | | | | | | | | | | | | | | | | | | | | |  
Db 27 GGCGAGGGAAGCGAA 12

RESULT 8  
AUI03768/c

LOCUS AUI03768 50 bp mRNA EST 05-APR-2001  
DEFINITION AUI03768 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
HEP12651, mRNA sequence.

ACCESSION AUI03768  
VERSION AUI03768.1 GI:13553289  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 50)  
Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata ,K., Suyama,A. and Sugano,S.  
Fine Structural analysis of transcription start sites of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL

COMMENT

Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yszukie@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano ,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1. .50  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP12651"

```

/clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT      8 a 14 c 11 g 17 t
ORIGIN

Query Match      64.0%; Score 12.8; DB 107; Length 50;
Best Local Similarity 87.5%; Pred. No. 2.4e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggccgcctgaagcgaa 16
    ||||| | |||||
Db 39 GCGCGAGGGAAGCGAA 24

RESULT 9
AUI03771/c
LOCUS      AUI03771 50 bp mRNA EST 05-APR-2001
DEFINITION AUI03771 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP13114, mRNA sequence.
ACCESSION  AUI03771
VERSION    AUI03771.1 GI:13553292
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 50)
AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            ,K., Suyama,A. and Sugano,S.
            H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
TITLE     Fine Structural analysis of transcription start sites of human
            mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL   Unpublished (2001)
COMMENT   Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: yszuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES   source
            1..50
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="HEP13114"
            /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT      6 a 14 c 12 g 18 t
ORIGIN

Query Match      64.0%; Score 12.8; DB 107; Length 50;
Best Local Similarity 87.5%; Pred. No. 2.4e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggccgcctgaagcgaa 16
    ||||| | |||||
Db 47 GCGCGAGGGAAGCGAA 32

RESULT 10
AUI03787/c
LOCUS      AUI03787 50 bp mRNA EST 05-APR-2001
DEFINITION AUI03787 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP15173, mRNA sequence.
ACCESSION  AUI03787
VERSION    AUI03787.1 GI:13553308
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 50)

```

```

AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            ,K., Suyama,A. and Sugano,S.
            Fine Structural analysis of transcription start sites of human
            mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL   Unpublished (2001)
COMMENT   Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: yszuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES   Location/Qualifiers
            1..50
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="HEP15173"
            /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT      6 a 13 c 11 g 20 t
ORIGIN

Query Match      64.0%; Score 12.8; DB 107; Length 50;
Best Local Similarity 87.5%; Pred. No. 2.4e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggccgcctgaagcgaa 16
    ||||| | |||||
Db 47 GCGCGAGGGAAGCGAA 32

RESULT 11
AUI03791/c
LOCUS      AUI03791 50 bp mRNA EST 05-APR-2001
DEFINITION AUI03791 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP16708, mRNA sequence.
ACCESSION  AUI03791
VERSION    AUI03791.1 GI:13553312
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 50)
AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            ,K., Suyama,A. and Sugano,S.
            Fine Structural analysis of transcription start sites of human
            mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL   Unpublished (2001)
COMMENT   Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: yszuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES   Location/Qualifiers
            1..50
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="HEP16708"
            /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT      6 a 13 c 13 g 18 t
ORIGIN

Query Match      64.0%; Score 12.8; DB 107; Length 50;
Best Local Similarity 87.5%; Pred. No. 2.4e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 ggcgcgtgaagcga 16  
 ||||| |||||  
 Db 43 GCGCCAGGAAGCGAA 28

## RESULT 12

AU103974/c

LOCUS AU103974 50 bp mRNA EST 05-APR-2001  
 DEFINITION AU103974 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 HEP00461, mRNA sequence.

ACCESSION AU103974

VERSION AU103974.1 GI:13553495

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.

TITLE Fine

STRUCTURAL ANALYSIS OF TRANSCRIPTION START SITES OF HUMAN  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL Unpublished (2001)

COMMENT Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1..50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HEP00461"

/clone\_lib="Sugano Homo sapiens cDNA library"

6 a 20 c 14 g 10 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 64.0%; Score 12.8; DB 107; Length 50;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ggcgcgtgaagcgaag 17

||||| |||||

Db 25 GCGCCGGAAGCGAGG 10

## RESULT 13

AU103975/c

LOCUS AU103975 50 bp mRNA EST 05-APR-2001  
 DEFINITION AU103975 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 HEP10657, mRNA sequence.

ACCESSION AU103975

VERSION AU103975.1 GI:13553496

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.

TITLE Fine

STRUCTURAL ANALYSIS OF TRANSCRIPTION START SITES OF HUMAN  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL Unpublished (2001)

COMMENT Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: ysuzuki@ims.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1..50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HEP10657"

/clone\_lib="Sugano Homo sapiens cDNA library"

4 a 21 c 9 g 16 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 64.0%; Score 12.8; DB 107; Length 50;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ggcgcgtgaagcgaag 17

||||| |||||

Db 46 GCGCCGGAAGCGAGG 31

## RESULT 14

AU103977/c

LOCUS AU103977 50 bp mRNA EST 05-APR-2001  
 DEFINITION AU103977 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 HEP17580, mRNA sequence.

ACCESSION AU103977

VERSION AU103977.1 GI:13553498

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.

TITLE Fine

STRUCTURAL ANALYSIS OF TRANSCRIPTION START SITES OF HUMAN  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

JOURNAL Unpublished (2001)

COMMENT Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1..50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HEP17580"

/clone\_lib="Sugano Homo sapiens cDNA library"

5 a 21 c 12 g 12 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 64.0%; Score 12.8; DB 107; Length 50;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ggcgcgtgaagcgaag 17

||||| |||||

Db 28 GCGCCGGAAGCGAGG 13

## RESULT 15

AU103980/c

LOCUS AU103980 50 bp mRNA EST 05-APR-2001  
DEFINITION AU103980 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
KAT01573, mRNA sequence.  
ACCESSION AU103980  
VERSION AU103980.1 GI:13553501  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
,K., Suyama,A. and Sugano,S.  
TITLE Fine structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
JOURNAL Unpublished (2001)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: ysuzuki@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
,S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  
FEATURES  
source  
1..50  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="KAT01573"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 6 a 20 c 14 g 10 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 107; Length 50;  
Best Local Similarity 87.5%; Pred. No. 2.4e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 gcgccgtgaagcgaag 17  
||||| ||||| |  
Db 25 GCGCCGGGAGCGGAGG 10

Search completed: October 2, 2001, 15:00:46  
Job time: 10815 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:37 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-6  
Perfect score: 20  
Sequence: 1 gaaactgcagaagcactga 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_bal: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_pil: \*  
13: gb\_p12: \*  
14: gb\_p13: \*  
15: gb\_p14: \*  
16: em\_bal: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rod: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rod: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pl: \*  
48: em\_ro: \*  
49: em\_sts: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_vi: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_vil: \*  
59: gb\_vil2: \*  
60: gb\_htg1: \*  
61: gb\_htg2: \*  
62: gb\_htg3: \*  
63: gb\_htg4: \*  
64: gb\_htg5: \*  
65: gb\_htg6: \*  
66: gb\_htg7: \*  
67: gb\_htg8: \*  
68: gb\_htg9: \*  
69: gb\_htg10: \*  
70: gb\_htg11: \*  
71: gb\_htg12: \*  
72: gb\_htg13: \*  
73: gb\_htg14: \*  
74: gb\_htg15: \*  
75: gb\_htg16: \*  
76: gb\_htg17: \*  
77: gb\_htg18: \*  
78: gb\_htg19: \*  
79: gb\_htg20: \*  
80: gb\_htg21: \*  
81: gb\_htg22: \*  
82: gb\_htg23: \*  
83: gb\_htg24: \*  
84: gb\_htg25: \*  
85: gb\_pr1: \*  
86: gb\_pr2: \*  
87: gb\_pr3: \*  
88: gb\_pr4: \*  
89: gb\_pr5: \*  
90: gb\_pr6: \*  
91: gb\_pr7: \*  
92: gb\_pr8: \*  
93: gb\_pr9: \*  
94: gb\_rol: \*  
95: gb\_rol2: \*  
96: gb\_in4: \*  
97: gb\_pr10: \*  
98: em\_ba3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	13.8	69.0	25	10 I26640	I26640 Sequence 3
c 2	13.8	69.0	42	95 S77087	S77087 T-cell rece
c 3	13.6	68.0	38	91 EPPROTII2	AF195645 Erythroce
4	13.6	68.0	38	97 SEPROTII2	AF195643 Semnopath
5	13.4	67.0	36	9 A69037	A69037 Sequence 25
c 6	13.4	67.0	43	9 AR018055	AR018055 Sequence
c 7	13.4	67.0	43	9 AR071945	AR071945 Sequence
8	13.4	67.0	44	10 E05905	E05905 Primer. 9/1

```
9 13.2 66.0 27 10 I27338 Sequence 13
10 13 65.0 44 10 E04313 PCR primer
11 12.8 64.0 22 9 A45762 Sequence 19
12 12.8 64.0 22 9 AR069418 Sequence 1
13 12.8 64.0 22 10 I43148
14 12.8 64.0 28 9 AR08097
15 12.8 64.0 28 9 AR08101
16 12.8 64.0 33 9 AR08100
17 12.8 64.0 34 9 AG2490
18 12.8 64.0 35 9 A69043
19 12.8 64.0 35 9 HUMTCRVJ55
20 12.6 63.0 21 9 AR011280
21 12.6 63.0 21 10 I17918
22 12.6 63.0 30 9 AR066368
23 12.6 63.0 34 9 S80826
24 12.4 62.0 25 9 AR003624
25 12.4 62.0 33 9 AR08096
26 12.4 62.0 39 10 I79231
27 12.2 61.0 18 9 AR101074
28 12.2 61.0 18 10 AX068306
29 12.2 61.0 27 10 I33779
30 12.2 61.0 30 10 AX080135
31 12.2 61.0 31 9 AR078949
32 12.2 61.0 42 9 A36503
33 12.2 61.0 42 9 AR080136
34 12.2 61.0 42 10 AX097533
35 12.2 61.0 50 9 AR032784
36 12.2 61.0 50 10 I29524
37 12.2 61.0 50 10 I91198
38 12 60.0 23 9 A84872
39 12 60.0 23 9 AR089998
40 12 60.0 29 9 A03947
41 12 60.0 29 9 A15005
42 12 60.0 30 9 A46142
43 12 60.0 30 10 I79830
44 12 60.0 32 9 AR052426
45 12 60.0 32 9 AR082414
```

## ALIGNMENTS

```
RESULT 1
126640/c I26640 25 bp DNA PAT 07-OCT-1996
LOCUS Sequence 3 from patent US 5559009.
DEFINITION I26640
ACCESSION I26640
VERSION I26640.1 GI:1606510
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 25)
AUTHORS Chandy,K.G., Kalman,K., Chandy,G. and Gutman,G.A.
TITLE Voltage-gated potassium channel gene, KV1.7, vectors and host cells comprising the same, and recombinant methods of making potassium channel proteins
JOURNAL Patent: US 5559009-A 3 24-SEP-1996;
FEATURES Location/Qualifiers
source 1..25
/organism="unknown"
BASE COUNT 2 a 11 c 5 g 7 t
ORIGIN
```

```
Query Match 69.0%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 2.le+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 gaactgcagaaggcac 17
| | | | | | | | | | | | | | | | |
Db 17 GGAAGTCAGAGGAC 1
```

```
RESULT 2
S77087/c
LOCUS S77087 42 bp mRNA ROD 14-AUG-2000
DEFINITION T-cell receptor beta chain V-J region [CDR3 region] [mice, transgenic, mRNA Partial, 42 nt].
ACCESSION S77087
VERSION S77087.1 GI:242717
KEYWORDS
SOURCE Mus sp. transgenic.
ORGANISM Mus sp.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 42)
AUTHORS Jorgensen,J.L., Esser,U., de St Groth,B.F., Reay,P.A. and Davis,M.M.
TITLE Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics
JOURNAL Nature 355 (6357), 224-230 (1992)
MEDLINE 92114961
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 77087] from the original journal article. This sequence comes from FIG. 4 C.
FEATURES
source 1..42
location/Qualifiers
/organism="Mus sp."
/db_xref="taxon:10095"
<1..>42
/note="This sequence comes from FIG. 4 c; conceptual translation presented here differs from translation in publication"
/codon_start=1
/product="T-cell receptor beta chain V-J region"
/protein_id="AAB20964.2"
/db_xref="GI:9800659"
/translation="CASSAFWGVAEQFF"
BASE COUNT 5 a 10 c 13 g 14 t
ORIGIN
```

```
Query Match 69.0%; Score 13.8; DB 95; Length 42;
Best Local Similarity 88.2%; Pred. No. 2.le+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 3 aactgcagaaggcacgtg 19
| | | | | | | | | | | | | | | | |
Db 25 AACCCAGAGGCACTG 9
```

```
RESULT 3
EPPROTII2
LOCUS EPPROTII2 38 bp DNA PRI 23-DEC-2000
DEFINITION Erythrocebus patas protamine P2 gene, exon 2 and complete cds.
ACCESSION AF195645
VERSION AF195645.1 GI:11990583
KEYWORDS
SEGMENT 2 of 2
SOURCE red quonon.
ORGANISM Erythrocebus patas
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopithecinae; Erythrocebus. 1 (bases 1 to 38)
AUTHORS Karanth,P., Stewart,C.B., Holt,R.A., deKoning,J. and Messier,W.
TITLE Positive Darwinian selection on the lineage leading to humans
JOURNAL Unpublished
AUTHORS Karanth,P. and Stewart,C.B.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1999) Biological Sciences, State University of New York, 1400 Washington Avenue, Albany, NY 12222, USA
FEATURES Location/Qualifiers
source 1..38
```



```

/organism="Erythrocebus patas"
/db_xref="taxon:9538"
/notes="common name:Patas monkey"
join(AF195644.1.<1..274,1..>38)
/product="protamine P2"
join(AF195644.1.1..274,1..38)
/codon_start=1
/product="protamine P2"
/protein_id="AAG42203.1"
/db_xref="GI:11990584"
/translation="MYRYRSLRSPHEVHGQOVYQEGHNGQEQGLSPHVEVY
ERTHQGSHRRRCRQRRLRIHRRRHSRRRRRSCRRRRRRRRCRTRRRRCR
Y"
exon 1..>38
/number=2
BASE COUNT 15 a 6 c 13 g 4 t
ORIGIN
Query Match 68.0%; Score 13.6; DB 91; Length 38;
Best Local Similarity 80.0%; Pred. No. 2.7e+04;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaaactgcagaagcactga 20
   ||| ||||| ||||| |||
Db 19 GAAGATGCAGAAGGTACTAA 38

RESULT 4
SEPROTI12
LOCUS SEPOTI12 38 bp DNA PRI 23-DEC-2000
DEFINITION Semnopithecus entellus protamine 2 gene, exon 2 and complete cds.
ACCESSION AF195643
VERSION AF195643.1 GI:11990579
KEYWORDS
SEGMENT 2 of 2
SOURCE Semnopithecus entellus.
ORGANISM Semnopithecus entellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
Colobinae; Semnopithecus.
REFERENCE 1 (bases 1 to 38)
AUTHORS Karanth,P., Stewart,C.B., Holt,R.A., deKoning,J. and Messier,W.
TITLE Positive Darwinian selection on the lineage leading to humans
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 38)
AUTHORS Karanth,P. and Stewart,C.B.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1999) Biological Sciences, State University of
New York, 1400 Washington Avenue, Albany, NY 12222, USA
FEATURES
source 1..38
/organism="Semnopithecus entellus"
/db_xref="taxon:88029"
/notes="common name:Hanuman langur"
join(AF195642.1.<1..274,1..>38)
/product="protamine 2"
join(AF195642.1.1..274,1..38)
/codon_start=1
/product="protamine 2"
/protein_id="AAG42202.1"
/db_xref="GI:11990580"
/translation="MYRYRMSLSRSPHEVHGQOVYQEGHNGQEQGLSPHVEVY
ERTHQGSHRRRCRQRRLRIHRRRHSRRRRRSCRRRRRRRRCRTRRRRCR
Y"
exon 1..>38
/number=2
BASE COUNT 15 a 6 c 13 g 4 t
ORIGIN
Query Match 68.0%; Score 13.5; DB 97; Length 38;
Best Local Similarity 80.0%; Pred. No. 2.7e+04;

```

```

Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaaactgcagaagcactga 20
   ||| ||||| ||||| |||
Db 19 GAAGATGCAGAAGGTACTAA 38

RESULT 5
A69037
LOCUS A69037 36 bp DNA PAT 06-MAY-1999
DEFINITION Sequence 25 from Patent WO9803660.
ACCESSION A69037
VERSION A69037.1 GI:4759966
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 36)
AUTHORS Baudu,P., Riviere,M., Audonnet,J. and Bouchardon,A.
TITLE FELINE POLYNUCLEOTIDE VACCINE FORMULA
JOURNAL Patent: WO 9803660-A 25 29-JAN-1998;
BAUDU PHILIPPE (FR)
COMMENT Other publication FR 2751223 19980123.
FEATURES
source 1..36
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 14 a 6 c 11 g 5 t
ORIGIN
Query Match 67.0%; Score 13.4; DB 9; Length 36;
Best Local Similarity 93.3%; Pred. No. 3.4e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aaactgcagaagcga 16
   ||||| ||||| |||
Db 2 AAATGCAGAAGGAA 16

RESULT 6
AR018055/c
LOCUS AR018055 43 bp DNA PAT 05-DEC-1998
DEFINITION Sequence 12 from patent US 5780286.
ACCESSION AR018055
VERSION AR018055.1 GI:3973658
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 43)
AUTHORS Dillon,P.J. and Vockley,J.G.
TITLE Arginase II
JOURNAL Patent: US 5780286-A 12 14-JUL-1998;
FEATURES
source 1..43
Location/Qualifiers
/organism="unknown"
BASE COUNT 8 a 10 c 8 g 17 t
ORIGIN
Query Match 67.0%; Score 13.4; DB 9; Length 43;
Best Local Similarity 93.3%; Pred. No. 3.4e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaagcact 18
   ||||| ||||| |||
Db 16 ACTGCAGAAGGCAAT 2

RESULT 7
AR071945/c
LOCUS AR071945 43 bp DNA PAT 18-FEB-2000

```

DEFINITION Sequence 12 from patent US 5912159.  
ACCESSION AR071945  
VERSION AR071945.1 GI:7222833  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 43)  
AUTHORS Vockley,J.G and Dillon,P.J.  
TITLE Arginase II  
JOURNAL Patent: US 5912159-A 12 15-JUN-1999;  
FEATURES Location/Qualifiers  
1..43  
/organism="unknown"  
BASE COUNT 8 a 10 c 8 g 17 t  
ORIGIN

Query Match 67.0%; Score 13.4; DB 9; Length 43;  
Best Local Similarity 93.3%; Pred. No. 3.4e+04;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 actgcagaagcact 18  
|||||  
Db 16 ACTGCAGAAGCAAT 2

RESULT 8  
E05905 E05905 44 bp DNA PAT 29-SEP-1997  
LOCUS  
DEFINITION Primer.  
ACCESSION E05905  
VERSION E05905.1 GI:2174092  
KEYWORDS JP 1993308963-A/3.  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 44)  
AUTHORS Kimura,S., Ishikawa,H. and Nakamura,H..  
TITLE METHOD FOR STABILIZING PROTEIN MOLECULE  
JOURNAL Patent: JP 1993308963-A 3 22-NOV-1993;  
COMMENT TANPAKU KOGAKU KENKYUSHO:KK  
OS Artificial gene  
OC Artificial sequence; Genes.  
PN JP 1993308963-A/3  
PD 22-NOV-1993  
PF 08-MAY-1992 JP 1992115877  
PI KIMURA SHIGENOBU, ISHIKAWA HIROKI, NAKAMURA HARUKI PC  
PC C12N9/22,C07K3/08,C12N15/55,C12P21/00;  
CC strandedness: Single;  
CC topology: Linear;  
FEATURES Location/Qualifiers  
1..44  
/organism="synthetic construct"  
BASE COUNT 17 a 9 c 11 g 7 t  
ORIGIN

Query Match 67.0%; Score 13.4; DB 10; Length 44;  
Best Local Similarity 93.3%; Pred. No. 3.4e+04;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 aaactgcagaagcca 16  
|||||  
Db 4 AACTGCAGACGGCA 18

RESULT 9  
I27338 I27338 27 bp DNA PAT 06-FEB-1997  
LOCUS  
DEFINITION Sequence 13 from patent US 5565320.  
ACCESSION I27338

VERSION I27338.1 GI:1818114  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 27)  
AUTHORS Cimino,G.D. and Lin,L.  
TITLE Quality control assay for platelet decontamination  
JOURNAL Patent: US 5565320-A 13 15-OCT-1996;  
FEATURES Location/Qualifiers  
1..27  
/organism="unknown"  
BASE COUNT 10 a 7 c 6 g 4 t  
ORIGIN

Query Match 66.0%; Score 13.2; DB 10; Length 27;  
Best Local Similarity 83.3%; Pred. No. 4.4e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aactgcagaagcactga 20  
|||||  
Db 2 AACTGCAGAAGCTAGGA 19

RESULT 10  
E04313 E04313 44 bp DNA PAT 29-SEP-1997  
LOCUS  
DEFINITION PCR primer for introducing point mutation into Escherichia coli  
ACCESSION E04313  
VERSION E04313.1 GI:2172516  
KEYWORDS JP 1993038286-A/3.  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 44)  
AUTHORS Kimura,S..  
TITLE VARIANT TYPE ESCHERICHIA COLI RIBONUCLEASE H  
JOURNAL Patent: JP 1993038286-A 3 19-FEB-1993;  
COMMENT TANPAKU KOGAKU KENKYUSHO:KK  
OS Artificial gene  
OC Artificial sequence; Genes.  
PN JP 1993038286-A/3  
PD 19-FEB-1993  
PF 07-AUG-1991 JP 1991197703  
PI KIMURA SHIGENOBU  
PC C12N9/22,C12N1/21,C12N15/55,(C12N9/22,C12R1:19),(C12N1/21, PC  
C12R1:19),  
PC (C12N15/55,C12R1:19);  
CC strandedness: Single;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No..  
FEATURES Location/Qualifiers  
1..44  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 15 a 10 c 12 g 7 t  
ORIGIN

Query Match 65.0%; Score 13; DB 10; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 actgcagaagcca 16  
|||||  
Db 6 ACTGCAGAAGGCA 18

RESULT 11  
A45762

LOCUS A45762 22 bp DNA 07-MAR-1997  
DEFINITION Sequence 19 from Patent WO9520672.  
ACCESSION A45762  
VERSION A45762.1 GI:2300149  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Cerutti,M., Chaabhi,H., Devauchelle,G., Gauthier,L., Kaczorek,M., Lefranc,M. and Poul,M.  
TITLE RECOMBINANT BACULOVIRUS AND USE THEREOF IN THE PRODUCTION OF MONOCLONAL ANTIBODIES  
JOURNAL Patent: WO 9520672-A 19 03-AUG-1995;  
COMMENT Other publication AU 1582495 950815  
Other publication FR 2715664 950804.  
FEATURES  
source 1. .22  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 5 a 3 c 7 g 3 t 4 others  
ORIGIN

Query Match 64.0%; Score 12.8; DB 9; Length 22;  
Best Local Similarity 68.8%; Pred. No. 7.1e+04;  
Matches 11; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 gaaactgcagaagcca 16  
:|||||:|:|:  
Db 5 SAAMCTGCAGRAGTCW 20

RESULT 12  
LOCUS AR069418 22 bp DNA 18-FEB-2000  
DEFINITION Sequence 1 from patent US 5891648.  
ACCESSION AR069418  
VERSION AR069418.1 GI:7220306  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Martin,M.T., Smith,R.G., Darsley,M.J., Simpson,D.M. and Blackburn,G.F.  
TITLE Reaction-based selection for expression of and concentration of catalytic moieties  
JOURNAL Patent: US 5891648-A 1 06-APR-1999;  
FEATURES  
source 1. .22  
/organism="unknown"  
BASE COUNT 7 a 3 c 9 g 3 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 9; Length 22;  
Best Local Similarity 87.5%; Pred. No. 7.1e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaaactgcagaagcca 16  
|||||:|:|:  
Db 5 GAAACTGCAGGAGTCA 20

RESULT 13  
LOCUS I43148 22 bp DNA 07-OCT-1997  
DEFINITION Sequence 1 from patent US 5631137.  
ACCESSION I43148  
VERSION I43148.1 GI:2468392  
KEYWORDS

SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Martin,M.T., Smith,R.G., Darsley,M.J., Simpson,D.M. and Blackburn,G.F.  
TITLE Reaction-based selection for expression of and concentration of catalytic moieties  
JOURNAL Patent: US 5631137-A 1 20-MAY-1997;  
FEATURES  
source 1. .22  
/organism="unknown"  
BASE COUNT 7 a 3 c 9 g 3 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 10; Length 22;  
Best Local Similarity 87.5%; Pred. No. 7.1e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaaactgcagaagcca 16  
|||||:|:|:  
Db 5 GAAACTGCAGGAGTCA 20

RESULT 14  
LOCUS A08097 28 bp DNA 20-JUL-1993  
DEFINITION Synthetic oligonucleotide M2.  
ACCESSION A08097  
VERSION A08097.1 GI:413338  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 28)  
AUTHORS Heinzel-Wieland,R., Ammann,J., Steffens,G.J. and Flohe,L.  
TITLE Serine protease inhibitor proteins, medicaments containing them, DNA sequences coding for these proteins and methods for producing these proteins, medicaments and DNA sequences  
JOURNAL Patent: EP 0373335-A 19 20-JUN-1990;  
FEATURES  
source 1. .28  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 6 a 8 c 10 g 4 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 9; Length 28;  
Best Local Similarity 87.5%; Pred. No. 7.1e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ctgcagaagcactga 20  
|||||:|:|:  
Db 7 CAGCAGCAGGACTGA 22

RESULT 15  
LOCUS A08101 28 bp DNA 20-JUL-1993  
DEFINITION Synthetic oligonucleotide M6.  
ACCESSION A08101  
VERSION A08101.1 GI:413342  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 28)  
AUTHORS Heinzel-Wieland,R., Ammann,J., Steffens,G.J. and Flohe,L.  
TITLE Serine protease inhibitor proteins, medicaments containing them, DNA sequences coding for these proteins and methods for producing

these proteins, medicaments and DNA sequences  
Patent: EP 0373335-A 23 20-JUN-1990;

Chinese proteins, medicaments and DNA  
Patent: EP 0373335-A 23 20-JUN-1990;  
Gruenthal GmbH

## FEATURES

source	1. .28
--------	--------

```

1. 1.20
source
/organism="synthetic construct"
/db_xref="taxon:32630"

```

BASE COUNT	5 a	9 c	9 g	5 t
00000000	00000000	00000000	00000000	00000000

ORIGIN

```
Query Match      64.0%; Score 12.8; DB 9; Length 28;
Best Local Similarity 87.5%; Pred. No. 7.1e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 5 ctqcaqaaqqcactqa 20

УДК 62-50

7 CAGCAGCAGCAGCTCA 22

Search completed: October 2, 2001, 15:56:39  
Job time: 14163 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:35 ; Search time 876.95 Seconds  
(without alignments)  
14.320 Million cell updates/sec

Title: US-09-757-100B-6

Perfect score: 20

Sequence: 1 gaaactgcagaagcactga 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	22	Human focal adhesi
2	15	75.0	15	22	Human focal adhesi
3	14.8	74.0	40	21	Dystrophin gene sp
4	14	70.0	33	21	Neisseria species
5	13.8	69.0	25	16	Splice site for mo
6	13.8	69.0	40	17	Soluble mouse Fas
7	13.6	68.0	42	21	HIV Vpr C-terminal
8	13.6	68.0	42	21	HIV-1 fragment PCR
9	13.4	67.0	20	20	PCR primer used to
10	13.4	67.0	36	19	Primer AB067 for F
11	13.4	67.0	43	18	PCR primer 5R used

c	12	13.4	67.0	43	18	AAV00587	PCR primer 5R used
c	13	13.4	67.0	43	19	AAV17634	Homo sapiens argin
c	14	13.4	67.0	43	20	AAV77316	Arginase II cDNA c
c	15	13.4	67.0	43	21	AAV08070	Human arginase II
c	16	13.4	67.0	44	14	AAQ04649	Ribonuclease prime
c	17	13.4	67.0	44	15	AAQ03496	pJk95N Primer 3.
c	18	13.4	67.0	44	16	AAQ96275	E.coli variant rib
c	19	13.2	66.0	27	14	AAQ47315	PCR primer for mit
c	20	13.2	66.0	27	17	AAV47908	platelet mitochond
c	21	13.2	66.0	27	21	AAV58555	HSV-2 antigen UL50
c	22	13	65.0	44	14	AAQ38711	Mutagenic PCR prim
c	23	13	65.0	44	14	AAQ40640	Ribonuclease H mut
c	24	13	65.0	44	20	AAQ34253	phoA-luc construct
c	25	12.8	64.0	21	16	AAQ9581	Primer B7 (Group 4
c	26	12.8	64.0	22	14	AAQ47805	Mouse spleen RNA 5
c	27	12.8	64.0	35	19	AAV49413	Primer AB017 for C
c	28	12.8	64.0	35	22	AAQ90404	Flounder growth ho
c	29	12.8	64.0	36	20	AAQ56735	WO915652 primer 2
c	30	12.8	64.0	36	21	AAQ55237	Rat meglin PCR pri
c	31	12.6	63.0	21	13	AAQ35615	env primer HIV3B10
c	32	12.6	63.0	21	14	AAQ35338	PCR primer HIV3B10
c	33	12.6	63.0	25	20	AAQ10964	PCR primer for HIV
c	34	12.6	63.0	26	16	AAQ01157	HIV gp160 PCR sens
c	35	12.6	63.0	26	18	AAQ85543	gp160 PCR sense pr
c	36	12.6	63.0	26	19	AAV21769	HIV gp160 gene amp
c	37	12.6	63.0	30	20	AAV08912	PCR primer for SIV
c	38	12.6	63.0	36	21	AAQ95630	HIV-1 env gene nat
c	39	12.6	63.0	39	21	AAQ55133	Neisseria species
c	40	12.6	63.0	43	21	AAV72024	HIV-1 proviral DNA
c	41	12.4	62.0	25	18	AAV47491	Human Inducible nI
c	42	12.4	62.0	28	21	AAV14276	PCR primer pGBD-SM
c	43	12.4	62.0	29	19	AAV34689	Yeast YCF1 gene PC
c	44	12.4	62.0	30	21	AAV30159	PCR primer TL F2 u
c	45	12.4	62.0	35	21	AAV59858	Human apo C-III pr

ALIGNMENTS

- RESULT 1
- AAQ65538
- ID AAC65538 standard; DNA; 20 BP.
- XX AAC65538;
- XX 12-FEB-2001 (first entry)
- XX Human focal adhesion kinase antisense sequence #4.
- XX Human; focal adhesion kinase; FAK; signal transduction; cancer;
- XX embryonic development disorder; angiogenic disorder; wound healing;
- XX antisense; phosphothioate; ss.
- XX Homo sapiens.
- XX US6133031-A.
- XX 17-OCT-2000.
- XX 19-AUG-1999; 99US-0377310.
- XX 19-AUG-1999; 99US-0377310.
- XX (ISIS-) ISIS PHARM INC.
- XX Monia BP, Gaarde WA;
- XX Soluble mouse Fas
- XX WPI; 2001-006141/01.
- XX New antisense compounds for inhibiting focal adhesion kinase
- XX expression, especially useful for inhibiting retinal
- XX neovascularization, or for diagnosing and treating e.g. colon cancer -

PS Claim 3; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, and is implicated  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

XX Sequence 20 BP; 8 A; 4 C; 6 G; 2 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gaaactgcagaaggcactga 20  
|||||  
Db 1 gaaactgcagaaggcactga 20

RESULT 2

AAC65558  
ID AAC65558 standard; DNA; 15 BP.

XX AAC65558;

DT 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #24.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.

XX Homo sapiens.

XX US6133031-A.

PD 17-OCT-2000.

XX 19-AUG-1999; 99US-0377310.

XX 19-AUG-1999; 99US-0377310.

XX (ISIS-) ISIS PHARM INC.

XX Monia BP, Gaarde WA;

XX WPI; 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase

PT expression, especially useful for inhibiting retinal

PT neovascularization, or for diagnosing and treating e.g. colon cancer -

XX Example 2; Column 25; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, and is implicated  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

XX Sequence 15 BP; 6 A; 4 C; 4 G; 1 T; 0 other;

Query Match 75.0%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 aactgcagaaggcac 17  
|||||  
Db 1 aactgcagaaggcac 15

RESULT 3

AAZ48582/C

ID AAZ48582 standard; DNA; 40 BP.

XX AAZ48582;

XX 31-MAR-2000 (first entry)

XX Dystrophin gene specific primer R5.

XX Muscular dystrophy; rod domain; adeno-associated virus; AAV;

KW dystrophin gene; truncated; PCR primer; ss.

XX Homo sapiens.

XX JP11318467-A.

XX 24-NOV-1999.

XX 08-MAY-1998; 98JP-0142134.

XX 08-MAY-1998; 98JP-0142134.

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX (KOKU-) KOKURITSU SEISHIN SHINKAI CENT SOCHO.

XX WPI; 2000-100771/09.

XX A gene encoding a shortened dystrophin - useful for the treatment of  
PT muscular dystrophy

XX Example 1; Page 10; 4pp; Japanese.

XX The invention provides a gene for the treatment of muscular dystrophy  
CC having at least one rod repeat structure of hinge 1, hinge 4 and rod  
CC domain of dystrophin gene and having a base sequence of 4.5 kb. The gene  
CC and a gene-introducing medium consisting of an adeno-associated virus  
CC (AAV) vector or lentivirus vector containing the rod shortened  
CC dystrophin genes can be used for the genetic treatment of muscular  
CC dystrophy of low immune reaction. Sequences AAZ48572-584 represent  
CC dystrophin gene specific primers.

XX Sequence 40 BP; 4 A; 13 C; 12 G; 11 T; 0 other;

Query Match 74.0%; Score 14.8; DB 21; Length 40;  
Best Local Similarity 88.9%; Pred. No. 4.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aaactgcagaaggcactg 19  
|||||  
Db 21 AGACTGCAGAGGCCCTG 4

RESULT 4

AAZ54675

ID AAZ54675 standard; DNA; 33 BP.

XX AAZ54675;

XX 21-MAR-2000 (first entry)

XX Neisseria species ORF cloning PCR primer #60.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

XX antibacterial; gene therapy; PCR primer; ss.

26-JAN-1995: 95US-0378507.

PR 27-MAY-1994; 94US-0250478.

PA (COLS ) UNIV COLORADO.

PI Bellgrau D, Duke RC;

XX WPI; 1996-030252/03.

DR Use of Fas ligand - for suppressing lymphocyte-mediated immune responses, e.g. transplant rejection or auto-immune conditions

PS Disclosure; Page 32; 51pp; English.

XX This 5' DNA primer is used for the synthesis of the full-length soluble mouse Fas ligand gene. It is used in conjunction with the 3' DNA primer (AA07687).

XX Sequence 40 BP; 9 A; 13 C; 11 G; 7 T; 0 other;

Query Match 69.0%; Score 13.8; DB 17; Length 40;

Best Local Similarity 88.2%; Pred. No. 1.3e+03;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 aactgcagaaggaactg 19

Db | ||||| ||||| |||||

RESULT 7

AAZ88079/c

ID AAZ88079 standard; DNA; 42 BP.

XX AC AAZ88079;

XX 20-APR-2000 (first entry)

DE HIV Vpr C-terminal and Tat N-terminal PCR primer SEQ ID NO:15.

XX Lentiviral vector; packaging; gag; pol; gene therapy; infection;

KW gene expression; PCR primer; ss.

XX Human immunodeficiency virus type 1.

OS WO200000600-A2.

XX 06-JAN-2000.

XX 26-MAY-1999; 99WO-US11516.

XX 26-MAY-1998; 98US-0086635.

XX (CHAN/) CHANG L.

XX Chang L;

XX WPI; 2000-137067/12.

XX New packaging vector comprising a nucleotide sequence encoding Gag and Pol proteins of a reference lentivirus useful for the delivery of non-lentiviral genes to target cells

XX Example 1; Page 148; 311pp; English.

XX The present invention describes a packaging vector (PV) comprising a nucleotide sequence encoding Gag and Pol proteins of a reference lentivirus that differs from the reference lentivirus at least in that: (a) its major splice donor site is either deleted or is sufficiently different from the reference lentivirus so that it is not a potential site for homologous recombination; and (b) it lacks a functional major packaging signal so that the introduced vector causes the host cell to produce packaging vector particles comprising functional Gag and Pol proteins. The vectors are useful for transforming (eukaryotic) cells to

CC express specific genes at high levels, e.g. for gene therapy. The improved vectors are safer, yet permit increased efficiency of packaging the recombinant viral genome and increased long-term gene expression. These properties are required for gene therapy as a means of treating infectious and non-infectious diseases. Unlike other retroviruses, the lentiviruses are able to infect non-dividing cells. The present sequence represents an HIV Vpr C-terminal and Tat N-terminal PCR primer which is used in the exemplification of the present invention.

XX Sequence 42 BP; 4 A; 15 C; 5 G; 18 T; 0 other;

Query Match 68.0%; Score 13.6; DB 21; Length 42;

Best Local Similarity 80.0%; Pred. No. 1.6e+03;

Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gaaactgcagaaggaactga 20

Db | |||| | |||| | ||||

RESULT 8

AAZ56558/G

ID AAZ56558 standard; DNA; 42 BP.

XX AC AAZ56558;

XX 21-MAR-2000 (first entry)

XX HIV-1 fragment PCR primer #2.

DE Lentiviral vector; gene amplification system; gene therapy; packaging; expression; transgenic animal; ss.

XX Human immunodeficiency virus type 1.

OS WO9961598-A2.

XX 02-DEC-1999.

XX 26-MAY-1999; 99WO-US11634.

XX 26-MAY-1998; 98US-0086635.

XX (UVFL ) UNIV FLORIDA.

XX Chang L, Cui Y, Iwakuma T;

XX WPI; 2000-072617/06.

XX New gene amplification system, useful to express a target gene in cells e.g. in gene therapy

XX Example 3; Page 98; 197pp; English.

XX The present invention describes a gene amplification system comprising a transducing vector, a packaging vector and a suitable host. The lentiviral vectors are useful to transform suitable host cells to express a target gene at high levels in the cells. For example, the vector system can be used therapeutically with both dividing and non-dividing human cells (e.g. neuronal cells and haematopoietic stem cells) e.g. in gene therapy. The vector system is also useful to produce polypeptides in cells (e.g. TE671 and HeLa cells) in vitro and in vivo, and to produce transgenic animals e.g. animals expressing human proteins. Use of modified lentiviral packaging vectors reduces the risk of generating replication-competent virus through recombination with the transducing vector or a defective provirus endogenous to the host cell, compared with prior art lentivirus packaging vectors. The vector system provides increased long-term gene expression compared with other viral vectors (e.g. could remain transduced for at least 120 days), so it may be possible for the vector to remain transduced for the necessary time course of a particular treatment. AAZ56538 to AAZ54578 represent nucleotide sequences used in the exemplification of



CC the present invention.

XX Sequence 42 BP; 4 A; 15 C; 5 G; 18 T; 0 other;

SQ Query Match 68.0%; Score 13.6; DB 21; Length 42;  
Best Local Similarity 80.0%; Pred. No. 1.6e+03;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gaaactgcagaagcactga 20  
||||| ||||| |||  
DB 42 GAAAGAGCAGAAGACAGTGA 23

RESULT 9  
AAZ04825/C  
ID AAZ04825 standard; DNA; 20 BP.  
XX AAZ04825;

AC 07-OCT-1999 (first entry)  
DE PCR primer used to amplify an ORF of Chlamydia trachomatis.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis; PCR primer;  
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.

XX Synthetic.  
OS Chlamydia trachomatis.

XX WO928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-1B01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST ) GENSET.

XX Griffais R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Disclosure; Page 1720; 1755pp; English.

XX PCR primers AAZ01426-206209 were used to amplify open reading frames  
CC (ORFs) of the genome of Chlamydia trachomatis (see AAZ01425). These ORFs  
CC encode polypeptides (see AAY36754-Y37949) which can be used as vaccines  
CC against Chlamydia trachomatis. Antisense and ribozyme sequences  
CC can also be used to control growth of the microorganism. Chlamydia  
CC trachomatis is responsible for a large number of diseases, e.g. eye  
CC diseases such as conventional trachoma, nonendemic trachoma,  
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
CC perihhepatitis, bartholinitis; pneumopathy in breast feeding infants;  
CC and venereal lymphogranulomatosis. The polypeptides of the  
CC invention may be of use in treating these diseases.

XX Sequence 20 BP; 5 A; 5 C; 5 G; 5 T; 0 other;

OY Query Match 67.0%; Score 13.4; DB 20; Length 20;  
Best Local Similarity 93.3%; Pred. No. 1.9e+03;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 actgcagaagcact 18

DB 19 ACTGCAGAACTCACT 5  
||||| |||||

RESULT 10  
AAV49285  
ID AAV49285 standard; DNA; 36 BP.

XX AAV49285;

XX 28-OCT-1998 (first entry)

XX Primer AB067 for FIV env gene.

XX Multivalent vaccine; cat; pathogen; respiratory disease; FeLV; FIV; PCV;  
KW digestive disease; feline leukaemia virus; feline panleukopaenia virus;  
KW feline calicivirus; feline immunodeficiency virus; FIV; rabies virus;  
KW vector; envelope glycoprotein; primer; PCR; amplification; ss.

XX Synthetic.

OS Feline immunodeficiency virus.

XX FR2751223-A1.

XX 23-JAN-1998.

XX 19-JUL-1996; 96FR-0009337.

XX 19-JUL-1996; 96FR-0009337.

XX (INMR ) RHONE MERIEUX SA.

XX Audonnet JCF, Baudu P, Bouchardon A, Riviere MEA;

XX WPI; 1998-112823/11.

XX Multi-valent polynucleotide vaccines against feline pathogens -  
PT consist of at least 3 plasmids able to express protective antigens  
PT from specified viruses

XX Example 17; Page 16; 42pp; French.

XX The invention relates to a multivalent vaccine for protecting cats  
CC against several pathogens, especially pathogens associated with  
CC respiratory and digestive diseases. The pathogens are especially  
CC selected from feline leukaemia virus (FeLV), feline panleukopaenia  
CC virus (FPV), feline calicivirus (FCV), feline immunodeficiency virus  
CC (FIV), or rabies virus. The vaccines are preferably composed of  
CC polynucleotide sequences encoding 3 antigens, all as part of vectors.  
CC Primers AAV49285-V49286 were used to PCR amplify the feline  
CC immunodeficiency virus (FIV) Petaluma strain gene encoding the envelope  
CC glycoprotein. The sequence sub-cloned into the plasmid pVR1012 to  
CC generate plasmid pAB030 for use in the vaccine.

XX Sequence 36 BP; 14 A; 6 C; 11 G; 5 T; 0 other;

OY Query Match 67.0%; Score 13.4; DB 19; Length 36;  
Best Local Similarity 93.3%; Pred. No. 2e+03;

XX Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 aaactgcagaaggca 16  
||||| |||||

DB 2 aaactgcagaaggaa 16

RESULT 11  
AAT84971/C  
ID AAT84971 standard; DNA; 43 BP.

XX AAT84971;

XX 01-APR-1998 (first entry)

```

XX DE PCR primer 5R used to amplify arginase II from a cDNA library.
XX DE
XX DE Arginase II; proline production; glutamate production; hyperarginaemia;
XX DE nitric oxide biosynthesis; arginase activity; urea cycle disease;
XX KW hypertension; hypotension; hyperammonaemia; prostate disease;
XX KW PCR primer; ss.
XX OS Synthetic.
XX PN WO9733985-A1.
XX PD 18-SEP-1997.
XX PF 14-MAR-1996; 96WO-US03561.
XX PR 14-MAR-1996; 96WO-US03561.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIK) SMITHKLINE BEECHAM CORP.
XX PI Dillon PJ, Vockley JG;
XX DR WPI; 1997-470867/43.
XX PT Polynucleotide encoding human arginase II - useful to treat,
XX PT diagnose and monitor, e.g. urea cycle disorders, prostatic disease,
XX PT hypotension and nitric oxide mediated immune and nervous diseases
XX PS Example 1; Fig 3; 85pp; English.
XX PS
XX CC PCR primers AAT84962-74 were used to amplify human arginase II from a
XX CC Jurkat cell line cDNA library. The cDNA sequence of Arginase I was used
XX CC as a probe sequence for a computer search of cDNA databases. Several
XX CC expressed sequence tags with 50-60% sequence homology were identified.
XX CC These were combined to give a 1075 bp sequence. The above PCR primers
XX CC were designed to the extreme 3' and 5' ends of the consensus sequence,
XX CC and used to isolate arginase II. In addition to a hypothetical role in
XX CC the production of proline and glutamate, it is postulated that
XX CC arginase II may play an important role in nitric oxide biosynthesis
XX CC through the production of ornithine as a precursor of glutamate.
XX CC Arginase II, or its agonists, antagonists and fragments, are used to
XX CC treat conditions associated with lack of arginase activity. Compounds
XX CC that inhibit activation of the protein are used to treat conditions
XX CC associated with excess arginase activity. Typical conditions that can be
XX CC treated are diseases of the urea cycle, hypertension, hypotension
XX CC (caused by sepsis or cytokines), episodic hyperammonaemia, defective
XX CC synthesis of proline, glutamate, nitric oxide or ornithine,
XX CC hyperarginaemia and related spasticity, prostate disease (e.g. cancer,
XX CC prostatitis and benign hypertrophy), prostate or kidney damage, also
XX CC nitric oxide associated immune and nervous system diseases. The arginase
XX CC II cDNA is used to produce recombinant protein and for chromosome
XX CC identification, while its fragments are used (as primers and probes) to
XX CC detect arginase II-encoding sequences and to diagnose the above
XX CC diseases.
XX SQ Sequence 43 BP; 8 A; 10 C; 8 G; 17 T; 0 other;

Query Match 67.0%; Score 13.4; DB 18; Length 43;
Best Local Similarity 93.3%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaaggcact 18
Db 16 ACTGCAGAAGGCAAT 2

RESULT 12
AAV00587/c
XX ID AAV00587 standard; cDNA; 43 BP.
XX AC AAV00587;

```

```

XX DT 25-MAR-1998 (first entry)
XX DE
XX DE PCR primer 5R used to amplify arginase II from a cDNA library.
XX DE
XX DE Arginase II; proline production; glutamate production; hyperarginaemia;
XX DE nitric oxide biosynthesis; arginase activity; urea cycle disease;
XX KW hypertension; hypotension; hyperammonaemia; prostate disease;
XX KW PCR primer; ss.
XX OS Synthetic.
XX PN WO9733986-A1.
XX PD 18-SEP-1997.
XX PF 20-AUG-1996; 96WO-US13455.
XX PR 14-MAR-1996; 96WO-US03561.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIK) SMITHKLINE BEECHAM CORP.
XX PI Dillon PJ, Vockley JG;
XX DR WPI; 1997-470868/43.
XX PT Nucleic acid encoding human arginase II - useful for treating,
XX PT diagnosing and monitoring e.g. urea cycle disorders, hypotension,
XX PT nitric oxide-mediated immune and nervous diseases, etc
XX PS Example 1; Fig 3; 93pp; English.
XX PS
XX CC PCR primers AAV00578-90 were used to amplify human arginase II from a
XX CC Jurkat cell line cDNA library. The cDNA sequence of Arginase I was used
XX CC as a probe sequence for a computer search of cDNA databases. Several
XX CC expressed sequence tags with 50-60% sequence homology were identified.
XX CC These were combined to give a 1075 bp sequence. The above PCR primers
XX CC were designed to the extreme 3' and 5' ends of the consensus sequence,
XX CC and used to isolate arginase II. In addition to a hypothetical role in
XX CC the production of proline and glutamate, it is postulated that
XX CC arginase II may play an important role in nitric oxide biosynthesis
XX CC through the production of ornithine as a precursor of glutamate.
XX CC Arginase II, or its agonists, antagonists and fragments, are used to
XX CC treat conditions associated with lack of arginase activity. Compounds
XX CC that inhibit activation of the protein are used to treat conditions
XX CC associated with excess arginase activity. Typical conditions that can be
XX CC treated are diseases of the urea cycle, hypertension, hypotension
XX CC (caused by sepsis or cytokines), episodic hyperammonaemia, defective
XX CC synthesis of proline, glutamate, nitric oxide or ornithine,
XX CC hyperarginaemia and related spasticity, prostate disease (e.g. cancer,
XX CC prostatitis and benign hypertrophy), prostate or kidney damage, also
XX CC nitric oxide associated immune and nervous system diseases. The arginase
XX CC II cDNA is used to produce recombinant protein and for chromosome
XX CC identification, while its fragments are used (as primers and probes) to
XX CC detect arginase II-encoding sequences and to diagnose the above
XX CC diseases.
XX SQ Sequence 43 BP; 8 A; 10 C; 8 G; 17 T; 0 other;

Query Match 67.0%; Score 13.4; DB 18; Length 43;
Best Local Similarity 93.3%; Pred. NO. 2e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaaggcact 18
Db 16 ACTGCAGAAGGCAAT 2

RESULT 13
AAV17634/c
XX ID AAV17634 standard; DNA; 43 BP.
XX ID AAV17634 standard; DNA; 43 BP.

```

```

XX AAV17634;
XX AC
XX DT
XX DE 03-AUG-1998 (first entry)
XX DE Homo sapiens arginase II gene PCR primer.
XX human; arginase II; disease; gene defect; hyperammonaemia;
XX diagnosis; susceptibility; urea cycle; hypertension; hypotension;
XX episodic; prolapse; blosynthesis; defect; glutamate; nitric oxide;
XX ornithine; hyperargininaemia; spasticity; growth retardation;
XX mental impairment; prostate; cancer; prostatic; benign prostatic;
XX hyperplasia; hypertrophy; damage; kidney; PCR primer; ss.
XX Synthetic.
XX OS Homo sapiens.
XX PN EP825260-A2.
XX PD
XX DE 25-FEB-1998.
XX PF
XX DE 20-AUG-1997; 97EP-0306360.
XX PR
XX DE 20-AUG-1996; 96US-0700186.
XX PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX PI
XX Dillion P, Vockley JG;
XX WPI; 1998-132255/13.
XX New human arginase II protein - useful for treating, e.g. diseases
XX associated with defect in arginase II gene such as episodic
XX hyperammonaemia
XX Example 1; Fig 3; 46pp; English.
XX The sequence is that of a PCR primer (5R) which was synthesised
XX based on a potential arginase II EST consensus sequence. It was
XX used in the amplification and cloning of the human arginase II gene.
XX Sequence 43 BP; 8 A; 10 C; 8 G; 17 T; 0 other;

Query Match 67.0%; Score 13.4; DB 19; Length 43;
Best Local Similarity 93.3%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaaggcact 18
DB 16 ACTGCAGAGGCAAT 2
|||||
RESULT 14
AAW77316/c
ID AAX77316 standard; DNA; 43 BP.
XX AC AAX77316;
XX DT
XX DE 09-AUG-1999 (first entry)
XX DE Arginase II cDNA cloning primer 5R.
XX Arginase II; gene therapy; hyperargininemia; immunogen; human;
XX urea cycle disorder; hypertension; hypotension; prostate cancer;
XX benign prostatic hyperplasia; PCR primer; ss.
XX Synthetic.
XX OS Homo sapiens.
XX PN US5912159-A.
XX

```

```

PD 15-JUN-1999.
XX XX
XX 20-AUG-1997; 97US-0914981.
XX PR
XX 20-AUG-1997; 97US-0914981.
XX PR
XX 20-AUG-1996; 96US-0700186.
XX XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX PI
XX Dillion PJ, Vockley JG;
XX WPI; 1999-357201/30.
XX Polypeptides with Arginase activity, useful for treating type I
XX arginase and disorders resulting in hyperargininemia
XX Examples; Fig 3; 38pp; English.
XX The invention relates to a polypeptide with Arginase II activity. The
XX Arginase II polypeptides may be used in gene therapy for the treatment
XX of type I arginase disorders resulting in hyperargininemia. The
XX polypeptide sequences may also be used as immunogens to produce
XX antibodies, these antibodies may then be used to treat numerous diseases
XX e.g. urea cycle disorders, hypertension, hypotension, prostate cancer,
XX benign prostatic hyperplasia. Sequences AAX77307-319 represent PCR
XX primers used for cloning the Arginase II cDNA.
XX Sequence 43 BP; 8 A; 10 C; 8 G; 17 T; 0 other;

Query Match 67.0%; Score 13.4; DB 20; Length 43;
Best Local Similarity 93.3%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaaggcact 18
DB 16 ACTGCAGAGGCAAT 2
|||||
RESULT 15
AAW08070/c
ID AAA08070 standard; DNA; 43 BP.
XX AC AAA08070;
XX DT 21-JUN-2000 (first entry)
XX DE Human arginase II PCR primer SEQ ID NO:12.
XX Human; arginase II; arginase I; diagnosis; hypotensive; hypertensive;
XX uropathic; cytostatic; neuroprotective; gene therapy; hypertension;
XX nitric oxide biosynthesis modulator; urea cycle disease; hypotension;
XX episodic hyperammonaemia; hyperargininaemia; spasticity; prostatic;
XX growth retardation; progressive mental impairment; prostate disease;
XX prostate cancer; benign prostatic hyperplasia; hypertrophy;
XX prostate damage; kidney disease; kidney damage; PCR primer; ss.
XX OS Homo sapiens.
XX PN US6054308-A.
XX PD 25-APR-2000.
XX PF 15-JUL-1998; 98US-0116115.
XX PR 14-MAR-1996; 96US-0013395.
XX PR 20-AUG-1997; 97US-0914981.
XX PR 20-AUG-1996; 96US-0700186.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX

```

PI Dillon PJ, Vockley JG;  
 XX WPI; 2000-328355/28.  
 XX  
 PT Novel human arginase II polypeptides useful for treating urea cycle  
 PT diseases, hypertension, hypotension, episodic hyperammonemia, to  
 PT control nitric oxide formation and kidney damage -  
 XX  
 XX Example 1; Fig 3; 37pp; English.  
 PS  
 XX The present invention describes human arginase II. Arginase II has  
 CC hypotensive, hypertensive, uropathic, cytostatic and neuroprotective  
 CC activities, and can be used in gene therapy and as a nitric oxide  
 CC biosynthesis modulator. Human arginase II proteins can be used to treat  
 CC diseases associated with or caused by a defect in the arginase II gene  
 CC or arginase II gene expression, such as, for e.g. urea cycle diseases,  
 CC hypertension, hypotension, episodic hyperammonemia, defects in  
 CC biosynthesis of proline, glutamate, nitric oxide and ornithine, as well  
 CC as hyperargininaemia and its related spasticity, growth retardation,  
 CC and progressive mental impairment, and prostate disease, particularly  
 CC prostate cancer, prostatitis and benign prostatic hyperplasia or  
 CC hypertrophy, and also prostate damage, kidney disease and kidney  
 CC damage. It is also used to control nitric oxide formation in an  
 CC individual. Arginase II or its fragments, variants or derivatives can be  
 CC used as diagnostic reagents for diagnosing arginase II deficiency in an  
 CC individual having or suspected of having a defect in the nitric oxide  
 CC pathway and the urea cycle. The genes encoding arginase II are used in  
 CC gene therapy techniques to treat the above mentioned disorders. It is  
 CC also used to deplete systemic arginine levels in an individual. AAA08061  
 CC to AAA08073 represent PCR primers used in the cloning of human arginase  
 CC II.  
 XX  
 SQ Sequence 43 BP; 8 A; 10 C; 8 G; 17 T; 0 other;

Query Match 67.0%; Score 13.4; DB 21; Length 43;  
 Best Local Similarity 93.3%; Pred. NO. 2e+03;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 actgcagaagcact 18  
 Db 16 ACTGCAGAGGCAAT 2  
 |||||

Search completed: October 2, 2001, 16:18:36  
 Job time: 15480 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:00:46 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-6  
Perfect score: 20  
Sequence: 1 gaaactgcagaagcactga 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues  
Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					TITLE	
Result No.	Score	Query Match %	Length	DB ID	Description	Journal Comment
C 1	12.8	64.0	33	250	AZ809697 2M0073M14	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
C 2	12.8	64.0	34	249	AZ759899 1M0553A10	Unpublished (2000)
C 3	12.6	64.0	37	8	AA519828 TgESTz235	Contact: Robert B. Weiss
C 4	12.6	63.0	37	139	BE739126 601556265	University of Utah
C 5	12.6	63.0	46	159	NA0673 yw78g11.r1	Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
C 6	12.6	63.0	47	244	AZ464547 1M0274A02	Tel: 801 585 5606
C 7	12.6	63.0	49	1	AA013617 mhl0005.r	Fax: 801 585 7177
C 8	12.6	63.0	50	107	AU102328 AU102328	Email: ddunn@genetics.utah.edu
C 9	12.4	62.0	50	107	AU102796 AU102796	Insert Length: 10000 Std Error: 0.00
C 10	12.4	62.0	50	107	AU102800 AU102800	Plate: 0073 row: M column: 14
C 11	12.2	61.0	31	249	AZ779631 2M0016K13	Seq primer: CACACAGCAACACGATGACG
C 12	12.2	61.0	38	243	AZ433216 1M0219A01	Class: plasmid ends
C 13	12.2	61.0	43	14	AA971669 op85C10.s	High quality sequence stop: 33.
C 14	12.2	61.0	43	15	AI048043 vn21f12.r	Location/Qualifiers
C 15	12.0	60.0	32	250	AZ826678 2M0102P06	1. .33
C 16	12.0	60.0	43	188	T17570 mps v30.the	/organism="Mus musculus"
C 17	12.0	60.0	46	11	AA781752 al60902.s	/strain="C57BL/6J"
C 18	12.0	60.0	48	249	AZ794791 2M0048B13	/db_xref="taxon:10090"
C 19	12.0	60.0	49	190	W77746 zd69a04.r1	/clone="UUGC2M0073M14"
C 20	12.0	60.0	49	246	AZ613610 1M0442J05	/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
C 21	12.0	60.0	50	107	AU106736 AU106736	/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
C 22	11.8	59.0	47	246	AZ582195 1M0374L22	(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114[gb]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
C 23	11.6	58.0	34	6	AA393427 zt75f10.r	BASE COUNT 7 a 9 c 5 g 12 t
C 24	11.6	58.0	43	8	AA509550 vH50D03.r	ORIGIN
C 25	11.6	58.0	45	250	AZ837415 2M0132K08	Query Match 64.0%; Score 12.8; DB 250; Length 33;
C 26	11.6	58.0	48	247	AZ634499 1M0490P16	Best Local Similarity 87.5%; Pred. No. 6.6e+04;
C 27	11.4	57.0	39	246	AZ592208 1M0402N22	Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
C 28	11.4	57.0	40	258	AL474254 T. brucei	
C 29	11.4	57.0	45	250	AZ817198 2M0086H20	
C 30	11.4	57.0	49	14	AA995077 ou53h10.s	
C 31	11.2	56.0	20	243	AZ434383 1M0220C17	
C 32	11.2	56.0	26	246	AZ591777 1M0402N04	
C 33	11.2	56.0	28	13	AA883279 aj14d10.s	
C 34	11.2	56.0	31	113	AW249392 2819331.5	
C 35	11.2	56.0	43	188	R83676 yp16f09.si	
C 36	11.2	56.0	43	246	AZ587450 1M0395D08	
C 37	11.2	56.0	49	103	AL877304 vz73e01.r	
C 38	11.2	56.0	49	258	TA197G04Q	
C 39	11.2	56.0	50	107	AU102905 AU102905	
C 40	11.2	56.0	50	107	AU102907 AU102907	
C 41	11.2	56.0	50	107	AU103044 AU103044	
C 42	11.2	56.0	50	107	AU103970 AU103970	
C 43	11.2	56.0	50	107	AU107111 AU107111	
C 44	11.2	56.0	50	107	AU107112 AU107112	
C 45	11.2	56.0	50	111	AW099485 sd41c08.y	
ALIGNMENTS					RESULTS	
					Qy	5 ctgcagaaggcactga 20
					Db	18 CAGCATAAGGCACTGA 3
					RESULT 2	
					LOCUS	AZ759899/c
					DEFINITION	1M0553A10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0553A10 F, DNA sequence.
					ACCESSION	AZ759899
					VERSION	AZ759899.1
					KEYWORDS	GI:12867159
					SOURCE	GSS.
					ORGANISM	house mouse.
					REFERENCE	Mus musculus
					AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 33)
						Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.





DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1C4493 row: a column: 17  
 High quality sequence stop: 33.  
 Location/Qualifiers

# FEATURES

```

source
1. .33
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3826072"
/clone_lib="NIH_MGC_58"
/tissue_type="hyponephroma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgccctggcc); Site_2: SfiI (ggcgccattggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-dt(30)BN-3'
(where B = A, C, G or N = A, C, G, or T). Average
insert size 1.35 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
long-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT      11 a      7 c      7 g      8 t
ORIGIN

```

Query Match 63.0%; Score 12.6; DB 139; Length 33;  
 Best Local Similarity 78.9%; Pred. No. 8.2e+04;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 aaactcagaagcactga 20
||||| ||||| |||||
Db 1 AACCTGTAGAAGACCCCTGA 19

```

```

RESULT 5
LOCUS N40673 46 bp mRNA EST 22-JAN-1996
DEFINITION Yw78g11.r1 Soares.Placenta_8to9weeks_2NBHP8to9W Homo sapiens cDNA
clone IMAGE:258404 5' similar to gb:J00117 CHORIOGONADOTROPIN BETA
CHAIN PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION N40673
VERSION N40673.1 GI:1164270
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 46)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,F., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Willson,R.
The WashU-Merck EST Project
Unpublished (1995)
JOURNAL Contact: Wilson RK
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Seq primer: T7
High quality sequence stop: 1.
Location/Qualifiers

```

# FEATURES

```

source
1. .46
/organism="Homo sapiens"
/db_xref="GDB:3888014"
/db_xref="taxon:9606"
/clone="IMAGE:258404"
/clone_lib="Soares.Placenta_8to9weeks_2NBHP8to9W"
/dev_stage="two placentae; one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Patima Bonaldo."
BASE COUNT      9 a      9 c      11 t      1 others
ORIGIN

```

# source

```

1. .46
/organism="Homo sapiens"
/db_xref="GDB:3888014"
/db_xref="taxon:9606"
/clone="IMAGE:258404"
/clone_lib="Soares.Placenta_8to9weeks_2NBHP8to9W"
/dev_stage="two placentae; one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Patima Bonaldo."
BASE COUNT      9 a      9 c      11 t      1 others
ORIGIN

```

Query Match 63.0%; Score 12.6; DB 159; Length 46;  
 Best Local Similarity 78.9%; Pred. No. 8.4e+04;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 gaaactcagaagcactg 19
||||| ||||| |||||
Db 45 GCAACAGCAGAAGCCCTG 27

```

```

RESULT 6
LOCUS AZ464547/c 47 bp DNA GSS 04-OCT-2000
DEFINITION IM0274A02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0274A02 F, DNA sequence.
ACCESSION AZ464547
VERSION AZ464547.1 GI:10622672
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 47)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
JOURNAL Contact: Robert B. Weiss
COMMENT University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0274 row: A column: 02
Seq primer: CGTTGTAACACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 47.
Location/Qualifiers

```

# FEATURES

```

source
1. .47
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0274A02"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"

```

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gii4732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 10 a 9 c 7 g 21 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 244; Length 47;  
Best Local Similarity 78.9%; Pred. No. 8.4e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 aaactgcagaagcactga 20  
||||| ||||| ||||| |||||  
Db 25 AAACCTACTGAGACATGTA 7

RESULT 7  
LOCUS AA013617 49 bp mRNA EST 21-JAN-1997  
DEFINITION mh10d05.r1 Soares mouse placenta 4NDMP13.5 14.5 Mus musculus cDNA clone IMAGE:442089 5' similar to WP:F35G12.9 CE00978 ;, mRNA sequence.

ACCESSION AA013617  
VERSION AA013617.1 GI:1474645  
KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 49)  
REFERENCE Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisels, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Willson, R. and Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)

JOURNAL Contact: Marra M/Mouse EST Project  
COMMENT Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:267425

Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand

Seq primer: ETPRimer  
High quality sequence stop: 1.

FEATURES Location/Qualifiers  
source 1..49  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:442089"

/clone\_lib="Soares mouse placenta 4NDMP13.5 14.5"  
/sex="unknown"  
/tissue\_type="placenta"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: placenta; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5',  
TGTTACCAACTGTAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 7 a 13 c 15 g 14 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 1; Length 49;  
Best Local Similarity 78.9%; Pred. No. 8.5e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gaactgcagaagcactg 19  
|| ||||| ||||| |||||  
Db 22 GACCTGCAGATGGCAGAC 4

RESULT 8  
LOCUS AU102328/c

DEFINITION AU102328 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone KAT06002, mRNA sequence.

ACCESSION AU102328  
VERSION AU102328.1 GI:13551848  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo, K., Sugama, A. and Sugano, S.  
Fine Structural analysis of transcription start sites of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries Unpublished (2001)

JOURNAL Contact: Yutaka Suzuki  
COMMENT Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: [ysuzuki@ims.u-tokyo.ac.jp](mailto:ysuzuki@ims.u-tokyo.ac.jp)  
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Sugama, A. and Sugano, S.  
Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES Location/Qualifiers  
source 1..50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="KAT06002"  
/clone\_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 2 a 18 c 15 g 15 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 107; Length 50;  
Best Local Similarity 78.9%; Pred. No. 8.5e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 aaactgcagaagcactga 20  
||| ||||| ||||| |||||  
Db 40 AACCAGCAGAGCCCTGA 22

```

RESULT 9
AUI02796
LOCUS      50 bp      mRNA      EST      05-APR-2001
DEFINITION Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            CAS11037, mRNA sequence.
ACCESSION  AUI02796
VERSION     AUI02796.1 GI:13552317
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 50)
AUTHORS     Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            K., Suyama,A. and Sugano,S.
TITLE       Fine Structural analysis of transcription start sites of human
            mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL     Unpublished (2001)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: yszukie@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES             Location/Qualifiers
     source           1..50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CAS11037"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT          19 a      5 c      21 g      5 t
ORIGIN
1 | | | | | | | | | |
4 actgcagaagggcac 17
30 AGTCAGAGGCGAC 43

Query Match      62.0%; Score 12.4; DB 107; Length 50;
Best Local Similarity 92.9%; Pred. No. 1.1e+05;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaagggcac 17
| | | | | | | | | |
30 AGTCAGAGGCGAC 43

RESULT 10
AUI02800
LOCUS      50 bp      mRNA      EST      05-APR-2001
DEFINITION Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            CAS11356, mRNA sequence.
ACCESSION  AUI02800
VERSION     AUI02800.1 GI:13552321
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 50)
AUTHORS     Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            K., Suyama,A. and Sugano,S.
TITLE       Fine Structural analysis of transcription start sites of human
            mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL     Unpublished (2001)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: yszukie@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            S. Construction and characterization of a full length-enriched and

```

```

a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES             Location/Qualifiers
     source           1..50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CAS11356"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT          20 a      8 c      16 g      6 t
ORIGIN
1 | | | | | | | | | |
4 actgcagaagggcac 17
19 AGTCAGAGGCGAC 32

Query Match      62.0%; Score 12.4; DB 107; Length 50;
Best Local Similarity 92.9%; Pred. No. 1.1e+05;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaagggcac 17
| | | | | | | | | |
19 AGTCAGAGGCGAC 32

RESULT 11
AZ779631/c
LOCUS      31 bp      DNA      GSS      16-FEB-2001
DEFINITION 2M0016K13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0016K13 F, DNA sequence.
ACCESSION  AZ779631
VERSION     AZ779631.1 GI:12910478
KEYWORDS   GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 31)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
            Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunne@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0016 row: K column: 13
            Seq primer: CGTTGTAAACGACGCGCAGT
            Class: plasmid ends
            High quality sequence stop: 31.
            Location/Qualifiers
             source      1..31
                       /organism="Mus musculus"
                       /strain="C57BL/6J"
                       /db_xref="taxon:10090"
                       /clone="UUGC2M0016K13"
                       /clone_lib="Mouse 10kb plasmid UUGC1M library"
                       /sex="Male"
                       /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                       /note="vector: PWD42nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel

```

electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

7 a 8 c 8 g 8 t

Query Match 61.0%; Score 12.2; DB 249; Length 31;  
Best Local Similarity 82.4%; Pred. No. 1.2e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 actgcagaagcactga 20  
||||| ||||| ||  
DB 23 ACTGCTGAAGCAACGA 7

## RESULT .12

AZ433216 38 bp DNA GSS 03-OCT-2000  
LOCUS 1M0219A01F Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
DEFINITION clone UUGCLM0219A01 F, DNA sequence.  
ACCESSION AZ433216  
VERSION AZ433216.1 GI:10557229  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 38)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0219 row: A column: 01  
Seq primer: CGTTGTAACGACGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 38.  
FEATURES  
source  
1. 38  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCLM0219A01"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

8 a 7 c 6 g 17 t

Query Match 61.0%; Score 12.2; DB 243; Length 38;  
Best Local Similarity 82.4%; Pred. No. 1.3e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 aactgcagaagcactg 19  
||||| ||||| |||||  
DB 13 ATCTCTAGAAGCCACTG 29

## RESULT 13

AA971669 43 bp mRNA EST 07-JUL-1998  
LOCUS Op95c10.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:1583634 3' similar to SW:PWP2\_HUMAN Q15269 PERIODIC  
TRYPTOPHAN PROTEIN 2 HOMOLOG. ;, mRNA sequence.  
ACCESSION AA971669  
VERSION AA971669.1 GI:3146959  
KEYWORDS EST.  
SOURCE human.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 43)  
REFERENCE NC-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
TITLE Unpublished (1997)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: csapbs-remail.nih.gov  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 429 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.  
FEATURES  
source  
1. 43  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1583634"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"

/a note="Organ: pooled; Vector: pF7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT  
ORIGIN

8 a 8 c 12 g 15 t

Query Match 61.0%; Score 12.2; DB 14; Length 43;  
Best Local Similarity 82.4%; Pred. No. 1.3e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 2 aaactgcagaagcact 18
   ||| ||||| |||||
Db 34 AAAATGCAGTACGCAC 18

RESULT 14
AI048043
LOCUS      43 bp      mRNA      EST      08-JUL-1998
DEFINITION vn21f12.r1 Knowles Soltter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:1021871 5' similar to WP:rl3H5.2 CE03653
            RETINAGE-BINDING LIKE PROTEIN ; , mRNA sequence.
ACCESSION  AI048043
VERSION     AI048043
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 43)
AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE      The WashU-HHMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicinep
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:572647
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
Seq primer: -40RP
High quality sequence stop: 1.
FEATURES   Location/Qualifiers
            1..43
                /organism="Mus musculus"
                /strain="B6D2 F1/J"
                /db_xref="taxon:10090"
                /clone="IMAGE:1021871"
                /clone_lib="Knowles Soltter mouse blastocyst B1"
                /tissue_type="blastocyst"
                /dev_stage="embryo (pre-implantation)"
                /lab_host="DH10B"
                /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally from mRNA prepared
                from 800 blastocysts. Primer: SalI(dT):
                5'-CGTCGACCGTCGACCGTGTGTGTGTGT-3'. cDNAs were
                cloned into the NotI/SalI sites of a pSPORT vector (Life
                Technologies). Two different size selections: B1 (larger
                inserts) and B3."
BASE COUNT 13 a 5 c 16 g 9 t
ORIGIN
Query Match 61.0%; Score 12.2; DB 15; Length 43;
Best Local Similarity 82.4%; Pred. No. 1.3e+05;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 actgcagaagcactga 20
   ||| ||||| |||||
Db 1 AGTGACGAGGCATTGA 17

RESULT 15
AZ826678

```

```

LOCUS      32 bp      DNA      GSS      20-FEB-2001
DEFINITION clone UUGC2M0102P06 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0102P06 R, DNA sequence.
ACCESSION  AZ826678
VERSION     AZ826678.1 GI:12996586
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 32)
AUTHORS    Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            and Wright,D., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL    Unpublished (2000)
COMMENT    Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0102 row: P column: 06
            Seq primer: CACAGGGAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 32.
FEATURES   Location/Qualifiers
            1..32
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC2M0102P06"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptored DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pWD42 (gll4732114|gb|AF129072.1), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptored mouse DNA was annealed to
                adaptored vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
BASE COUNT 8 a 8 c 11 g 5 t
ORIGIN
Query Match 60.0%; Score 12; DB 250; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgcagaagcca 16
   ||||| |||||
Db 16 CTGCAGAAGGCA 27

```

Search completed: October 2, 2001, 15:00:48  
Job time: 10817 sec

. . . . .

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:39 ; Search time 3399.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-7

Perfect score: 20

Sequence: 1 ttctcccttcggtatttatt 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_bal.\*

2: gb\_ba2.\*

3: gb\_ba3.\*

4: gb\_in1.\*

5: gb\_in2.\*

6: gb\_in3.\*

7: gb\_om.\*

8: gb\_ov.\*

9: gb\_pat1.\*

10: gb\_pat2.\*

11: gb\_ph.\*

12: gb\_pl1.\*

13: gb\_pl2.\*

14: gb\_pl3.\*

15: gb\_pl4.\*

16: em\_bal.\*

17: em\_ba2.\*

18: em\_fun.\*

19: em\_htgo\_hum.\*

20: em\_htgo\_inv.\*

21: em\_htgo\_rod.\*

22: em\_htg\_hum1.\*

23: em\_htg\_hum2.\*

24: em\_htg\_hum3.\*

25: em\_htg\_hum4.\*

26: em\_htg\_hum5.\*

27: em\_htg\_hum6.\*

28: em\_htg\_hum7.\*

29: em\_htg\_hum8.\*

30: em\_htg\_inv1.\*

31: em\_htg\_inv2.\*

32: em\_htg\_other.\*

33: em\_htg\_rod.\*

34: em\_hum1.\*

35: em\_hum2.\*

36: em\_hum3.\*

37: em\_hum4.\*

38: em\_hum5.\*

39: em\_hum6.\*

40: em\_hum7.\*

41: em\_in.\*

42: em\_om.\*

43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vi2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_ro1.\*  
95: gb\_ro2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	13.8	69.0	33	9 AX060862	AX060862 Sequence
2	13.8	69.0	41	9 AR109090	AR109090 Sequence
3	12.6	63.0	20	9 A81012	A81012 Sequence 64
4	12.6	63.0	20	9 A95391	A95391 Sequence 64
5	12.6	63.0	30	9 AR004745	AR004745 Sequence
6	12.6	63.0	30	9 AR008231	AR008231 Sequence
7	12.6	63.0	30	10 I77015	I77015 Sequence 75
8	12.6	63.0	30	10 I81010	I81010 Sequence 75

```

9      12.6 63.0 30 10 I81106 Sequence 75
10     12.2 61.0 22 20 AR10286 Sequence 4
11     12.2 61.0 22 10 I73726 Sequence 4
12     12.2 61.0 35 10 I03808 Sequence 11
13     12.2 60.0 26 9 AX043897 Sequence
14     12.2 60.0 26 9 AX059414 Sequence
15     12.2 60.0 27 9 AX043901 Sequence
16     12.2 60.0 30 9 A83876 Sequence 11
17     12.2 60.0 30 9 AR061353 Sequence
18     12.2 60.0 30 9 AR108252 Sequence
19     12.2 60.0 30 10 I16209 Sequence 35
20     12.2 60.0 30 10 I66695 Sequence 35
21     12.2 60.0 30 10 I84789 Sequence 35
22     12.2 60.0 31 9 A04979 Artificial
23     12.2 60.0 31 9 A04980 Artificial
24     12.2 60.0 34 9 AR036378 Sequence
25     12.2 60.0 34 10 I72126 Sequence 41
26     12.2 60.0 39 10 AX080565 Sequence
27     12.2 60.0 40 10 I15938 Sequence 39
28     12.2 60.0 40 10 I96137 Sequence 39
29     12.2 60.0 42 9 AR104422 Sequence
30     12.2 60.0 43 9 A46553 Sequence 9
31     12.2 60.0 45 9 A18117 oligonucleo
32     12.2 60.0 47 9 A82690 Sequence 35
33     12.2 60.0 47 9 A82705 Sequence 50
34     12.2 60.0 50 10 I34292 Sequence 6
35     11.8 59.0 23 9 A81700 Sequence 5
36     11.8 59.0 24 10 I51986 Sequence 20
37     11.8 59.0 27 9 AR039578 Sequence
38     11.8 59.0 45 9 AR089681 Sequence
39     11.6 58.0 18 9 AR030029 Sequence
40     11.6 58.0 19 9 A64833 Sequence 9
41     11.6 58.0 21 10 AX095134 Sequence
42     11.6 58.0 22 9 AR022302 Sequence
43     11.6 58.0 22 9 AR066394 Sequence
44     11.6 58.0 22 9 AR084122 Sequence
45     11.6 58.0 22 9 AX011523 Sequence

```

## ALIGNMENTS

```

RESULT 1
AX060862
LOCUS AX060862 33 bp DNA PAT 22-JAN-2001
DEFINITION Sequence 83 from Patent WO0078815.
ACCESSION AX060862
VERSION AX060862.1 GI:12406240
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 33).
AUTHORS Huse,W.D. and Wu,H.
TITLE Anti-5(a) v2-g(b)3? recombinant human antibodies, nucleic acids
JOURNAL encoding same and methods of use
PATENT: WO 0078815-A 83 28-DEC-2000;
APPLIED Molecular Evolution (US)
FEATURES
source
1..33
/organism="synthetic construct"
/db_xref="taxon:32630"
<1..>33
/note="Mutated complementarity determining region (CDR)"
CDS
/codon_start=1
/transl_table=11
/protein_id="CAC24917.1"
/db_xref="GI:12406241"
/translation="ILLIRYSQSSIS"
4 a 13 c 2 g 14 t
BASE COUNT
ORIGIN

```

```

Query Match 69.0%; Score 13.8; DB 9; Length 33;
Best Local Similarity 88.2%; Pred. No. 1.2e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ttcctccgttattctt 20
   ||| ||||| |||||
DB 3 TCTCATCCGTTATCTT 19
   ||| ||||| |||||

RESULT 2
ARI09090 ARI09090 41 bp DNA PAT 14-FEB-2001
LOCUS ARI09090 Sequence 49 from patent US 6114116.
DEFINITION
ACCESSION ARI09090
VERSION ARI09090.1 GI:12825366
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 41)
AUTHORS Lemieux,B., Landry,B.S. and Sapolsky,R.J.
TITLE Brassica polymorphisms
JOURNAL Patent: US 6114116-A 49 05-SEP-2000;
FEATURES
source
1..41
/organism="unknown"
BASE COUNT 8 a 11 c 3 g 18 t 1 others
ORIGIN

```

```

Query Match 69.0%; Score 13.8; DB 9; Length 41;
Best Local Similarity 78.9%; Pred. No. 1.2e+04;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttctcccttcgttattct 19
   ||||| ||| |||||
DB 12 TTCTCCAGTCTTATCT 30

```

```

RESULT 3
ARI012
LOCUS ARI012 20 bp DNA PAT 21-JAN-2000
DEFINITION Sequence 64 from Patent EP0918091.
ACCESSION ARI012
VERSION ARI012.1 GI:6731585
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 20)
AUTHORS Kahn,A. and Chelly,J.
TITLE A gene called XLIS and the XLIS gene product, called doublecortin
JOURNAL and their applications
PATENT: EP 0918091-A 64 26-MAY-1999;
INST NAT SANTE RECH MED (FR)
FEATURES
source
1..20
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1 a 7 c 0 g 12 t
ORIGIN

```

```

Query Match 63.0%; Score 12.6; DB 9; Length 20;
Best Local Similarity 78.9%; Pred. No. 4.9e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttctcccttcgttattct 19
   ||| ||||| |||
DB 2 TTATCCCTTCCTTCTCT 20

```



RESULT 4  
LOCUS A95391 20 bp DNA PAT 26-JAN-2000  
DEFINITION Sequence 64 from Patent WO9927089.  
ACCESSION A95391  
VERSION A95391.1 GI:6779435  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 20)  
AUTHORS Francis, F. and Kahn, A.  
TITLE A GENE CALLED XLIS AND THE XLIS GENE PRODUCT, CALLED DOUBLECORTIN AND THEIR PREPARATIONS  
JOURNAL Patent: WO 9927089-A 64 03-JUN-1999;  
INST NAT SANTE RECH MED (FR); FRANCIS FIONA (FR)  
FEATURES  
source Location/Qualifiers  
1..20  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1 a 7 c 12 t  
BASE COUNT 1 a 7 c 12 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 20;  
Best Local Similarity 78.9%; Pred. No. 4.9e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttctcccttcgcgttattct 19  
||||| ||||| || |||||  
Db 2 TTATCCCTTCCTCTCTCT 20

RESULT 5  
LOCUS AR004745 30 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 75 from patent US 5747282.  
ACCESSION AR004745  
VERSION AR004745.1 GI:3965624  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Skolnick, M.H., Goldgar, D.E., Miki, Y., Swenson, J., Kamb, A., Harshman, K.D., Shattuck-Eidens, D.M., Tavtigian, S.V., Wiseman, R.W. and Futreal, P. Andrew.  
TITLE 170-linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 5747282-A 75 05-MAY-1998;  
FEATURES  
source Location/Qualifiers  
1..30  
/organism="unknown"  
1 a 14 c 2 g 13 t  
BASE COUNT 1 a 14 c 2 g 13 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 30;  
Best Local Similarity 78.9%; Pred. No. 5e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttctcccttcgcgttattctt 20  
||||| ||||| || |||||  
Db 8 TCTCTCTCTCTCTCTCTT 26

RESULT 6  
LOCUS AR008231 30 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 75 from patent US 5753441.  
ACCESSION AR008231  
VERSION AR008231.1 GI:3967340  
KEYWORDS

Query Match 63.0%; Score 12.6; DB 9; Length 30;  
Best Local Similarity 78.9%; Pred. No. 5e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttctcccttcgcgttattctt 20  
||||| ||||| || |||||  
Db 8 TCTCTCTCTCTCTCTT 26

Query Match 63.0%; Score 12.6; DB 10; Length 30;  
Best Local Similarity 78.9%; Pred. No. 5e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttctcccttcgcgttattctt 20  
||||| ||||| || |||||  
Db 8 TCTCTCTCTCTCTCTT 26

Query Match 63.0%; Score 12.6; DB 9; Length 30;  
Best Local Similarity 78.9%; Pred. No. 5e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttctcccttcgcgttattctt 20  
||||| ||||| || |||||  
Db 8 TCTCTCTCTCTCTCTT 26

SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Skolnick, M.H., Goldgar, D.E., Miki, Y., Swenson, J., Kamb, A., Harshman, K.D., Shattuck-Eidens, D.M., Tavtigian, S.V., Wiseman, R.W. and Futreal, P. Andrew.  
TITLE 170-linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 5753441-A 75 19-MAY-1998;  
FEATURES  
source Location/Qualifiers  
1..30  
/organism="unknown"  
1 a 14 c 2 g 13 t  
BASE COUNT 1 a 14 c 2 g 13 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 30;  
Best Local Similarity 78.9%; Pred. No. 5e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttctcccttcgcgttattctt 20  
||||| ||||| || |||||  
Db 8 TCTCTCTCTCTCTCTT 26

RESULT 7  
LOCUS I77015 30 bp DNA PAT 03-APR-1998  
DEFINITION Sequence 75 from patent US 5693473.  
ACCESSION I77015  
VERSION I77015.1 GI:3013169  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Shattuck-Eidens, D.M., Simard, J., Durocher, F., Emi, M. and Nakamura, Y.  
TITLE Linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 5693473-A 75 02-DEC-1997;  
FEATURES  
source Location/Qualifiers  
1..30  
/organism="unknown"  
1 a 14 c 2 g 13 t  
BASE COUNT 1 a 14 c 2 g 13 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 30;  
Best Local Similarity 78.9%; Pred. No. 5e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttctcccttcgcgttattctt 20  
||||| ||||| || |||||  
Db 8 TCTCTCTCTCTCTCTT 26

RESULT 8  
LOCUS I81010 30 bp DNA PAT 10-JUN-1998  
DEFINITION Sequence 75 from patent US 5709999.  
ACCESSION I81010  
VERSION I81010.1 GI:3209300  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Shattuck-Eidens, D.M., Simard, J., Durocher, F., Emi, M. and Nakamura, Y.  
TITLE Linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 5709999-A 75 20-JAN-1998;  
FEATURES  
source Location/Qualifiers  
1..30

```

BASE COUNT      1 a      14 c      2 g      13 t
ORIGIN

Query Match      63.0%; Score 12.6; DB 10; Length 30;
Best Local Similarity 78.9%; Pred. No. 5e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy  2 tctccctccggttattc 20
    |||| ||||| | |||||
Db  8 TCTCTCTCTCTCTCTCT 26

RESULT 9
LOCUS      I81106      30 bp      DNA      PAT      10-JUN-1998
DEFINITION Sequence 75 from patent US 5710001.
ACCESSION  I81106
VERSION     I81106.1 GI:3209396
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 30)
AUTHORS   Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A.,
           Harshman,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W.
           and Futreal,P.Andrew.
TITLE     17q-linked breast and ovarian cancer susceptibility gene
JOURNAL   Patent: US 5710001-A 75 20-JAN-1998;
FEATURES   Location/Qualifiers
            1..30
            /organism="unknown"

BASE COUNT      1 a      14 c      2 g      13 t
ORIGIN

Query Match      63.0%; Score 12.6; DB 10; Length 30;
Best Local Similarity 78.9%; Pred. No. 5e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy  2 tctccctccggttattc 20
    |||| ||||| | |||||
Db  8 TCTCTCTCTCTCTCTCT 26

RESULT 10
LOCUS      AR110286      22 bp      DNA      PAT      14-FEB-2001
DEFINITION Sequence 38 from patent US 6114502.
ACCESSION  AR110286
VERSION     AR110286.1 GI:12826562
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 22)
AUTHORS   North,M., Nishina,P., Naggert,J. and Noben-Trauth,K.
TITLE     Gene family associated with neurosensory defects
JOURNAL   Patent: US 6114502-A 38 05-SEP-2000;
FEATURES   Location/Qualifiers
            1..22
            /organism="unknown"

BASE COUNT      2 a      10 c      1 g      9 t
ORIGIN

Query Match      61.0%; Score 12.2; DB 9; Length 22;
Best Local Similarity 82.4%; Pred. No. 7.8e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy  2 tctccctccggttattc 18
    |||| ||||| | |||

```

---

```

Db  2 TCTCCCTTCCTTCCTTC 18

RESULT 11
LOCUS      I73726      22 bp      DNA      PAT      03-APR-1998
DEFINITION Sequence 4 from patent US 5686598.
ACCESSION  I73726
VERSION     I73726.1 GI:3009867
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 22)
AUTHORS   North,M., Nishina,P. and Naggert,J.
TITLE     Genes associated with retinal dystrophies
JOURNAL   Patent: US 5686598-A 4 11-NOV-1997;
FEATURES   Location/Qualifiers
            1..22
            /organism="unknown"

BASE COUNT      2 a      10 c      1 g      9 t
ORIGIN

Query Match      61.0%; Score 12.2; DB 10; Length 22;
Best Local Similarity 82.4%; Pred. No. 7.8e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy  2 tctccctccggttattc 18
    |||| ||||| | |||
Db  2 TCTCCCTTCCTTCCTTC 18

RESULT 12
LOCUS      I03808      35 bp      PAT      02-DEC-1994
DEFINITION Sequence 11 from Patent EP 0055942.
ACCESSION  I03808
VERSION     I03808.1 GI:592013
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 35)
AUTHORS   Inouye,M. and Nakamura,K.
TITLE     Plasmid cloning vehicles
JOURNAL   Patent: EP 0055942-A2 11 14-JUL-1982;
FEATURES   Location/Qualifiers
            1..35
            /organism="unknown"

BASE COUNT      14 a      2 c      10 g      9 t
ORIGIN

Query Match      61.0%; Score 12.2; DB 10; Length 35;
Best Local Similarity 82.4%; Pred. No. 7.9e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy  1 ttctccctccggttatt 17
    ||| ||||| | |||||
Db  32 TTCCCCCTTCATTATT 16

RESULT 13
LOCUS      AX043897/c      26 bp      DNA      PAT      23-NOV-2000
DEFINITION Sequence 10 from Patent WO0049166.
ACCESSION  AX043897
VERSION     AX043897.1 GI:11342482
KEYWORDS   .
SOURCE     synthetic construct.
ORGANISM   synthetic construct
           artificial sequence.

```

REFERENCE 1 (bases 1 to 26)  
AUTHORS Graham,F.L., Ng,P., Parks,R., Bacchetti,S. and Anglana,M.  
TITLE A system for production of helper dependent adenovirus vectors based on use of endonucleases  
JOURNAL Patent: WO 0049166-A 10 24-AUG-2000;  
Merck & Co., Inc. (US)

FEATURES  
source  
1. .26  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 17 a 5 c 3 g 1 t  
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 26;  
Best Local Similarity 75.0%; Pred. No. 9.9e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
||||| ||| ||||| ||| |||  
Db 26 TTCTACTTTTCGTTTCTT 7

RESULT 14  
AX059414/c  
LOCUS AX059414 26 bp DNA PAT 17-JAN-2001  
DEFINITION Sequence 147 from Patent WO0055325.  
ACCESSION AX059414  
VERSION AX059414.1 GI:12311519  
KEYWORDS  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 11;  
Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 26)  
AUTHORS Preuss,D., Copenhaver,G. and Keith,K.  
TITLE Plant chromosome compositions and methods  
JOURNAL Patent: WO 0055325-A 147 21-SEP-2000;  
The University of Chicago (US)

FEATURES  
source  
1. .26  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"

BASE COUNT 13 a 1 c 10 g 2 t  
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 26;  
Best Local Similarity 75.0%; Pred. No. 9.9e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
||||| ||| ||||| ||| |||  
Db 20 TTCTCTCTCTCTCTCTCTT 1

RESULT 15  
AX043901/c  
LOCUS AX043901 27 bp DNA PAT 23-NOV-2000  
DEFINITION Sequence 14 from Patent WO0049166.  
ACCESSION AX043901  
VERSION AX043901.1 GI:11342484  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequence.

REFERENCE 1 (bases 1 to 27)  
AUTHORS Graham,F.L., Ng,P., Parks,R., Bacchetti,S. and Anglana,M.  
TITLE A system for production of helper dependent adenovirus vectors based on use of endonucleases  
JOURNAL Patent: WO 0049166-A 14 24-AUG-2000;

FEATURES  
source  
1. .27  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 18 a 5 c 3 g 1 t  
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 27;  
Best Local Similarity 75.0%; Pred. No. 9.9e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
||||| ||| ||||| ||| |||  
Db 27 TTCTACTTTTCGTTTCTT 8

Search completed: October 2, 2001, 15:56:41  
Job time: 14165 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:00:48 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-7

Perfect score: 20

Sequence: 1 ttctccctccgttattctt 20

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				JOURNAL	
Result No.	Score	Query Match	Length DB ID	COMMENT	
ALIGNMENTS				Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhls@sanger.ac.uk	
T. brucei sheared genomic DNA clone 124f02, reverse sequence, genomic survey sequence.				Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).	
Trypanosoma brucei.				Email: nelsayed@tigr.org	
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;				Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.	
Trypanosoma.				Location/Qualifiers	
1 (bases 1 to 35)				1. 35	
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.				/organism="Trypanosoma brucei"	
Direct Submission				/db_xref="taxon:5691"	
				/clone="124f02"	
				14 a 2 c 11 g 8 t	
				BASE COUNT	
				ORIGIN	
				Query Match 71.0%; Score 14.2; DB 258; Length 35;	
				Best Local Similarity 84.2%; Pred. No. 7.1e+03;	
				Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
				QY 2 tctccctccgttattctt 20	
				Db 28 TCTCCCTCCTGTTATTATT 10	
				RESULT 2	
				AZ810736	
				LOCUS	
				DEFINITION	
				ACCESSION	
				VERSION	
				KEYWORDS	
				SOURCE	
				ORGANISM	
				REFERENCE	
				AUTHORS	
				TITLE	
				JOURNAL	
				COMMENT	
				Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts	
				Unpublished (2000)	
				Contact: Robert B. Weiss	
				University of Utah Genome Center	
				University of Utah	
				Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA	
				Tel: 801 585 5606	
				Fax: 801 585 7177	
				Email: ddunn@genetics.utah.edu	
				Insert length: 10000 Std Error: 0.00	
				Plate: 0076 row: B column: 05	
				Seq primer: CACACAGAAACAGCTATGACC	
				Class: plasmid ends	
				High quality sequence stop: 49.	
				Location/Qualifiers	
				1. 49	
				/organism="Mus musculus"	
				/strain="C57BL/6J"	
				/db_xref="taxon:10090"	

```

/clone="UUGC2M0076B05"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114[gblAF129072.1]), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      4 a      13 c      0 g      32 t
ORIGIN

Query Match      68.0%; Score 13.6; DB 250; Length 49;
Best Local Similarity 80.0%; Pred. No. 1.4e+04;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20
    |||| |||| |||| ||||
Db 9 TTCTTCTTCATTATTCTT 28

RESULT 3
LOCUS      AA500031      37 bp      mRNA      EST      01-JUL-1997
DEFINITION clone IMAGE:920108 5' similar to SW:COX1_RAT P05503 CYTOCHROME C
            OXIDASE POLYPEPTIDE I ;, mRNA sequence.
ACCESSION  AA500031
VERSION     AA500031.1 GI:2234998
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 37)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE      The WashU-HMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:532324
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 1.
FEATURES
            Location/Qualifiers
                source
                    1..37

```

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:920108"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGAGCGCCGCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTTGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."
BASE COUNT      6 a      8 c      5 g      18 t
ORIGIN

Query Match      64.0%; Score 12.8; DB 8; Length 37;
Best Local Similarity 87.5%; Pred. No. 3.3e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtatt 16
    |||| |||| ||||
Db 2 TTCTCCCTTCATTAT 17

RESULT 4
LOCUS      AZ447206      30 bp      DNA      GSS      04-OCT-2000
DEFINITION lM0244J07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0244J07 F, DNA sequence.
ACCESSION  AZ447206
VERSION     AZ447206.1 GI:10598962
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 30)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL    Unpublished (2000)
COMMENT    Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0244 row: J column: 07
            Seq primer: CGTGTAAACGACGCCAGT
            Class: plasmid ends
            High quality sequence stop: 30.
FEATURES
            Location/Qualifiers
                source
                    1..30
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UUGC1M0244J07"
                    /clone_lib="Mouse 10kb plasmid UUGC1M library"
                    /sex="Male"
                    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                    /note="vector: PWD42nv; Purified genomic DNA from M.

```



musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMW42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 2 a 16 c 0 g 12 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 244; Length 30;  
Best Local Similarity 78.9%; Pred. No. 3.9e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tctcccttcggtattctt 19  
||||| ||||| | |||||  
Db 10 TTCTCTCTCTCTCTCT 28

## RESULT 5

AZ776590 50 bp DNA GSS 16-FEB-2001  
LOCUS  
DEFINITION 2M0010A23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0010A23 F, DNA sequence.

ACCESSION AZ776590  
VERSION  
KEYWORDS GSS.  
SOURCE

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 50)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhauser,A. and Wright,D.,Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

TITLE

JOURNAL

COMMENT

Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0010 row: A column: 23  
Seq primer: CGTTGTAACACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 50.  
Location/Qualifiers

FEATURES  
source

1. .50  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0010A23"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMW42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 22 c 1 g 24 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 249; Length 50;  
Best Local Similarity 78.9%; Pred. No. 4.3e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 tctcccttcggtattctt 20  
||||| ||||| | |||||  
Db 27 TTCTCTACCTTTTCTT 45

## RESULT 6

AZ404740 37 bp DNA GSS 03-OCT-2000  
LOCUS  
DEFINITION 1M0173D10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0173D10 F, DNA sequence.

ACCESSION AZ404740  
VERSION  
KEYWORDS GSS.  
SOURCE

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 37)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhauser,A. and Wright,D.,Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

TITLE

JOURNAL

COMMENT

Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0173 row: D column: 10  
Seq primer: CGTTGTAACACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 37.  
Location/Qualifiers

FEATURES  
source

1. .37  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0173D10"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 5 a 8 c 5 g 19 t  
 ORIGIN

Query Match 61.0%; Score 12.2; DB 243; Length 37;  
 Best Local Similarity 82.4%; Pred. No. 6.3e+04;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttctcccttcgatt 17  
 Db 20 TTCTCTTTTCGGTGATT 36

RESULT 7  
 AZ357911/c  
 LOCUS AZ357911 39 bp DNA GSS 02-OCT-2000  
 DEFINITION IM0099L13R Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0099L13 R, DNA sequence.

ACCESSION AZ357911  
 VERSION AZ357911.1 GI:10471611  
 KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 39)  
 REFERENCE Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0099 row: L column: 13  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 39.

FEATURES  
 source  
 1. .39  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0099L13"  
 /clone\_1lb="Mouse 10kb plasmid UUGCLM library"

/sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 17 a 0 c 16 g 6 t  
 ORIGIN

Query Match 61.0%; Score 12.2; DB 242; Length 39;  
 Best Local Similarity 82.4%; Pred. No. 6.4e+04;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttctcccttcgatt 17  
 Db 17 TTCTCCCTCCCATATT 1

RESULT 8

T49694

LOCUS

DEFINITION

T49694

46 bp mRNA

EST

06-FEB-1995

ya78g11.s1 Stratagene placenta (#937225) Homo sapiens cDNA clone

IMAGE:67844 3' similar to gb:M24194 GUANINE

NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN (HUMAN), mRNA

sequence.

ACCESSION T49694

VERSION T49694.1

GI:651554

EST

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 46)

REFERENCE Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

Chissoe, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W., Hawkins

, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore

, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,

Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,

Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

Other ESTs: ya78g11.r1

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Insert Size: 711

High quality sequence starts: 1

High quality sequence stops: 1

Source: IMAGE Consortium, LLNL This clone is available royalty-free

through LLNL; contact the IMAGE Consortium (info@image.llnl.gov)

for further information. Trace considered overall poor quality

Insert Length: 711 Std Error: 0.00

Seq primer: -21ml3

Seq quality sequence stop: 1.

## FEATURES

source  
1. .46  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:489509"  
/db\_xref="taxon:9606"  
/clone="IMAGE:67844"  
/clone\_lib="Stratagene placenta (#937225)"  
/sex="male"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: placenta; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Caucasian. Average insert size: 1.2 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'"  
5 a 13 c 6 g 18 t 4 others

BASE COUNT  
ORIGIN

Query Match 61.0%; Score 12.2; DB 189; Length 46;  
Best Local Similarity 73.7%; Pred. No. 6.5e+04;  
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttctccttcgttattct 19  
||||| ||||| ||| |||  
Db 26 TTCTGCTTCNGTTCNCT 44

## RESULT 9

AI747625 49 bp mRNA EST 22-JUN-1999  
LOCUS ul19d03.x1 sugano mouse embryo mewa Mus musculus cDNA clone  
DEFINITION IMAGE:2088005 3' similar to TR:089050 O89050 MUSKELIN. ; , mRNA  
sequence.  
ACCESSION AI747625  
VERSION AI747625.1 GI:5125827  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 49)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson.R.  
TITLE The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished (1999)  
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:995689

Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: custom primer used  
High quality sequence stop: 1.

## FEATURES

source  
1. .49  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2088005"  
/clone\_lib="Sugano mouse embryo mewa"  
/dev\_stage="embryo, 14 dpc"  
/lab\_host="DH10B"  
/note="Vector: pME18S-FL3; Site\_1: DraIII (CACTGTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTTTTTTTTTTTTT];

double-stranded cDNA was ligated to a DraIII adaptor [TCTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science).  
Custom primers for sequencing: 5' end primer CTCTCTCTCTAAAGCTGGG and 3' end primer CGACCTGCTAGCTGACACA." 22 t

BASE COUNT 11 a 11 c 5 g 22 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 24; Length 49;  
Best Local Similarity 82.4%; Pred. No. 6.6e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 tcctcttcgttattctt 20  
||||| ||||| ||| |||  
Db 28 TCCTCTCTCTTTTCTT 44

## RESULT 10

AZ385263/c 29 bp DNA GSS 02-OCT-2000  
LOCUS 1M0143109R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0143109 R, DNA sequence.

ACCESSION AZ385263  
VERSION AZ385263.1 GI:10498963  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 29)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

## TITLE

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
CONTACT: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu

## JOURNAL

## COMMENT

Insert Length: 10000 Std Error: 0.00  
Plate: 0143 row: 1 column: 09  
Seq primer: CACACGAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 29.

## FEATURES

## source

1. .29  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0143109"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114[gblAF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

```

BASE COUNT      12 a   3 c   12 g   2 t
ORIGIN

Query Match      60.0%; Score 12; DB 242; Length 29;
Best Local Similarity 75.0%; Pred. No. 7.6e+04;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttctccctccggtattctt 20
   ||||| || |||||
Db 27 TCCTCCCTCCCTTCCTT 8

RESULT 11
LOCUS      AI355496/c
DEFINITION qul5c07.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:1964844 3'
           similar to TR:Q60501 Q60501 P7 PREINSERTION DNA. ;contains element
           TARI repetitive element ; , mRNA sequence.
ACCESSION  AI355496
VERSION     AI355496.1 GI:4095649
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 31)
AUTHORS   Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
           Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
           M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
           and Wright, D., Weiss, R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
           plasmid inserts
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
           Emmert-Buck, M.D., Ph.D.
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 2334 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES
source      1..31
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1964844"
            /clone_lib="NCI_CGAP_Ov23"
            /tissue_type="tumor, 5 pooled (see description)"
            /lab_host="DH10B"
            /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 1.35 kb. Tumor types include: mixed
            Mullerian tumor, papillary serous, clear cell, spindle
            cell. All are primary tumors, metastasis positive. Life
            Technologies catalog #: 11534-013"

BASE COUNT      18 a   1 c   12 g   0 t
ORIGIN

```

```

Query Match      60.0%; Score 12; DB 19; Length 31;
Best Local Similarity 75.0%; Pred. No. 7.6e+04;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttctccctccggtattctt 20
   ||||| || |||||
Db 27 TCCTTCCTTCCTTCCTT 8

RESULT 12
LOCUS      AZ850974
DEFINITION 2M0153109F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
           clone UUGC2M0153109 F, DNA sequence.
ACCESSION  AZ850974
VERSION     AZ850974.1 GI:13036700
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 32)
AUTHORS   Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
           Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
           M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
           and Wright, D., Weiss, R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
           plasmid inserts
JOURNAL    Unpublished (2000)
COMMENT    Contact: Robert B. Weiss
           University of Utah Genome Center
           University of Utah
           Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
           84112, USA
           Tel: 801 585 5606
           Fax: 801 585 7177
           Email: ddunn@genetics.utah.edu
           Insert Length: 10000 Std Error: 0.00
           Plate: 0153 row: 1 column: 09
           Seq primer: CGTGTAAACGACGCGCCAGT
           Class: plasmid ends
           High quality sequence stop: 32.
FEATURES
source      1..32
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="UUGC2M0153109"
            /clone_lib="Mouse 10kb plasmid UUGC1M library"
            /sex="Male"
            /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
            /note="vector: PWD42nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adapted DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of pWD42 (g14732114[gblAF129072.1]), a copy-number
            inducible derivative of plasmid R1. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adapted mouse DNA was annealed to
            adapted vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."

BASE COUNT      13 a   0 c   17 g   2 t

```

## ORIGIN

Query Match 60.0%; Score 12; DB 250; Length 32;  
 Best Local Similarity 75.0%; Pred. No. 7.7e+04;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttctcccttcggtattctt 20  
 1 ||||| ||||| |||||  
 Db 24 TCCTCCCTACTTTCCTT 5

## RESULT 13

AI551509 34 bp mRNA EST 23-MAR-1999  
 LOCUS vo54h02.x1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA  
 DEFINITION clone IMAGE:1053747 3' similar to FR:Q33559 Q33559 NH2 TERMINUS  
 UNCERTAIN ; mRNA sequence.

ACCESSION AI551509  
 VERSION AI551509  
 KEYWORDS EST.

SOURCE AI551509.1 GI:4483872  
 house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 34)  
 AUTHOR Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person  
 B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter  
 E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.  
 TITLE The WashU-NCI Mouse EST Project 1999  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:585323  
 This clone was previously sequenced on the 5' end only, this new  
 data is from the 3' end  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 High quality sequence stop: 1.

## FEATURES

source Location/Qualifiers  
 1..34  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1053747"  
 /clone\_lib="Barstead mouse irradiated colon MPLRB7"  
 /dev\_stage="8 weeks"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained  
 from 8 week old mouse. Colon was harvested 72 hours after  
 irradiation with 1400 Gys. 1st strand cDNA was primed  
 with a Not I - oligo(dT) primer  
 [5'TGTACGAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors [AATCGATCCTTG], digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library constructed by Bob Barstead. "  
 4 a 9 c 1 g 20 t  
 BASE COUNT 4 a 9 c 1 g 20 t  
 ORIGIN

Query Match 60.0%; Score 12; DB 21; Length 34;  
 Best Local Similarity 75.0%; Pred. No. 7.7e+04;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttctcccttcggtattctt 20  
 1 ||||| ||||| |||||  
 Db 2 TTTTCTCTTATTATTCCT 21

## RESULT 14

AZ357613 34 bp DNA GSS 02-OCT-2000  
 LOCUS AZ357613/c Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 DEFINITION clone UUGCLM0099K18 F, DNA sequence.

ACCESSION AZ357613  
 VERSION AZ357613  
 KEYWORDS GSS.

SOURCE AZ357613.1 GI:10471313  
 house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 34)  
 AUTHOR Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,  
 and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0099 row: K column: 18  
 Seq primer: CGTTGTAACAGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 34.

## FEATURES

source Location/Qualifiers  
 1..34  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0099K18"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gil4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."  
 18 a 1 c 15 g 0 t  
 BASE COUNT 18 a 1 c 15 g 0 t  
 ORIGIN

Query Match 60.0%; Score 12; DB 242; Length 34;  
 Best Local Similarity 75.0%; Pred. No. 7.7e+04;

	Matches	15; Conservative	0; Mismatches	5; Indels	0; Gaps
Oy	1	ttctcccttcogttattctt	20		
Db	22	TCCTTCCTTTTCGTCTTCTT	3		

Best Local Similarity 75.0%; Pred. No. 7.8e+04;  
Matches 15; Conservative 0; Mismatches 5

QY	1	tctcccttcggtattctt	20
Db	32	tctctctctctctctctt	13

LOCUS	AZ829796	35 bp	DNA	GSS	20-FEB-2001
DEFINITION	2M0107H11R Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUC2M0107H11 R, DNA sequence.				

ACCESSION AZ829796  
 VERSION AZ829796.1 GI:12999704  
 KEYWORDS GSS.

REFERENCE 1 (bases 1 to 35)

**AUTHORS**  
Dunn, D., Ayoagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,  
and Wright, D., Weiss, R.

**TITLE**  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

**JOURNAL**  
**COMMENT**  
Unpublished (2000)  
Contact: Robert B. Weiss

University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177  
Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00

Seq primer: CACACAGGAAACAGCTATGACC

Class: plasmid ends  
High quality sequence stop: 35.

FEATURES	Location/Qualifiers
source	1. .35

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0107H11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource"

```

	BASE COUNT	ORIGIN
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000
9	0.000000	0.000000
10	0.000000	0.000000
11	0.000000	0.000000
12	0.000000	0.000000
13	0.000000	0.000000
14	0.000000	0.000000
15	0.000000	0.000000
16	0.000000	0.000000
17	0.000000	0.000000
18	0.000000	0.000000
19	0.000000	0.000000
20	0.000000	0.000000
21	0.000000	0.000000
22	0.000000	0.000000
23	0.000000	0.000000
24	0.000000	0.000000
25	0.000000	0.000000
26	0.000000	0.000000
27	0.000000	0.000000
28	0.000000	0.000000
29	0.000000	0.000000
30	0.000000	0.000000
31	0.000000	0.000000
32	0.000000	0.000000
33	0.000000	0.000000
34	0.000000	0.000000
35	0.000000	0.000000
36	0.000000	0.000000
37	0.000000	0.000000
38	0.000000	0.000000
39	0.000000	0.000000
40	0.000000	0.000000
41	0.000000	0.000000
42	0.000000	0.000000
43	0.000000	0.000000
44	0.000000	0.000000
45	0.000000	0.000000
46	0.000000	0.000000
47	0.000000	0.000000
48	0.000000	0.000000
49	0.000000	0.000000
50	0.000000	0.000000
51	0.000000	0.000000
52	0.000000	0.000000
53	0.000000	0.000000
54	0.000000	0.000000
55	0.000000	0.000000
56	0.000000	0.000000
57	0.000000	0.000000
58	0.000000	0.000000
59	0.000000	0.000000
60	0.000000	0.000000
61	0.000000	0.000000
62	0.000000	0.000000
63	0.000000	0.000000
64	0.000000	0.000000
65	0.000000	0.000000
66	0.000000	0.000000
67	0.000000	0.000000
68	0.000000	0.000000
69	0.000000	0.000000
70	0.000000	0.000000
71	0.000000	0.000000
72	0.000000	0.000000
73	0.000000	0.000000
74	0.000000	0.000000
75	0.000000	0.000000
76	0.000000	0.000000
77	0.000000	0.000000
78	0.000000	0.000000
79	0.000000	0.000000
80	0.000000	0.000000
81	0.000000	0.000000
82	0.000000	0.000000
83	0.000000	0.000000
84	0.000000	0.000000
85	0.000000	0.000000
86	0.000000	0.000000
87	0.000000	0.000000
88	0.000000	0.000000
89	0.000000	0.000000
90	0.000000	0.000000
91	0.000000	0.000000
92	0.000000	0.000000
93	0.000000	0.000000
94	0.000000	0.000000
95	0.000000	0.000000
96	0.000000	0.000000
97	0.000000	0.000000
98	0.000000	0.000000
99	0.000000	0.000000
100	0.000000	0.000000

Query Match	.	60.0%	Score 12;	DB 250;	Length 35;
-------------	---	-------	-----------	---------	------------

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:03:46 ; Search time 417.38 seconds  
(without alignments)  
9.071 Million cell updates/sec

Title: US-09-757-100B-7

Perfect score: 20  
Sequence: 1 ttctcccttcggtattctt 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 460742

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCRUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	20	3	US-09-377-310-7
2	15	75.0	15	3	US-09-377-310-27
3	13.8	69.0	41	3	US-08-813-507-49
4	12.6	63.0	30	1	US-08-480-784-75
5	12.6	63.0	30	1	US-08-483-553-75
6	12.6	63.0	30	1	US-08-487-002-75
7	12.6	63.0	30	1	US-08-483-554B-75
8	12.6	63.0	30	1	US-08-488-011B-75
9	12.6	63.0	30	4	PCT-US95-10202-75
10	12.6	63.0	30	5	PCT-US95-10203-75
11	12.6	63.0	30	5	PCT-US95-10207-75
12	12.6	63.0	30	5	PCT-US95-10220-75
13	12.2	61.0	22	1	US-08-701-380-4
14	12.2	61.0	22	3	US-09-032-365A-38
15	12	60.0	30	1	US-07-931-473B-35
16	12	60.0	30	1	US-07-714-131C-35
17	12	60.0	30	1	US-08-412-110-35
18	12	60.0	30	2	US-08-409-442A-35
19	12	60.0	30	3	US-09-143-190-35
20	12	60.0	34	1	US-08-413-813-41
21	12	60.0	34	2	US-08-467-346-41
22	12	60.0	38	3	US-08-881-094-34
23	12	60.0	39	2	US-08-308-887A-6
24	12	60.0	39	3	US-08-881-094-6
25	12	60.0	40	1	US-08-199-507B-39
26	12	60.0	40	1	US-08-441-828-39
27	12	60.0	40	1	US-08-441-828-39

c 28	12	60.0	42	3	US-08-879-565-6	Sequence 6, Appli
c 29	12	60.0	50	1	US-08-133-629-6	Sequence 6, Appli
c 30	11.8	59.0	24	1	US-08-470-179-201	Sequence 201, App
c 31	11.8	59.0	27	1	US-08-758-306-426	Sequence 426, App
c 32	11.8	59.0	45	2	US-08-823-516-158	Sequence 158, App
c 33	11.6	58.0	18	2	US-08-173-489C-218	Sequence 218, App
c 34	11.6	58.0	22	1	US-08-524-757-36	Sequence 36, Appl
c 35	11.6	58.0	22	2	US-08-457-273B-18	Sequence 18, Appl
c 36	11.6	58.0	22	2	US-08-955-138-25	Sequence 25, Appl
c 37	11.6	58.0	33	3	US-08-872-417B-2	Sequence 2, Appli
c 38	11.6	58.0	35	3	US-08-872-417B-1	Sequence 1, Appli
c 39	11.6	58.0	39	3	US-09-074-357-13	Sequence 13, Appl
c 40	11.6	58.0	41	2	US-08-472-719-8	Sequence 8, Appli
c 41	11.6	58.0	41	2	US-08-472-719-9	Sequence 9, Appli
c 42	11.6	58.0	43	1	US-08-473-096-3	Sequence 3, Appli
c 43	11.6	58.0	46	1	US-08-171-389-98	Sequence 98, Appl
c 44	11.6	58.0	46	1	US-08-123-936-98	Sequence 98, Appl
c 45	11.6	58.0	46	2	US-08-475-228A-98	Sequence 98, Appl

ALIGNMENTS

RESULT 1  
US-09-377-310-7  
; Sequence 7, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-7

Query Match 100.0%; Score 20; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
Db 1 ttctcccttcggtattctt 20

RESULT 2  
US-09-377-310-27  
; Sequence 27, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 27  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-27

Query Match 75.0%; Score 15; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ctcccttcggtatt 17

Db 1 ctcccttcggtatt 15

## RESULT 3

US-08-813-507-49  
; Sequence 49, Application US/08813507  
; Patent No. 611416

; GENERAL INFORMATION:  
; APPLICANT: Lemieux, Bertrand  
; APPLICANT: Landry, Benoit S.  
; APPLICANT: Sapolsky, Ronald J.  
; TITLE OF INVENTION: Brassica Polymorphisms  
; NUMBER OF SEQUENCES: 173

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/813.507  
; FILING DATE: 07-MAR-1997

; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/032,069  
; FILING DATE: 02-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Liebeschuetz, Joe  
; REGISTRATION NUMBER: 37,505  
; REFERENCE/DOCKET NUMBER: 018547-030100US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415 576-0200  
; TELEFAX: 415 576-0200

; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 41 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-813-507-49

Query Match 69.0%; Score 13.8; DB 3; Length 41;  
Best Local Similarity 78.9%; Pred. No. 2.4e+02;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattct 19

Db 12 ttctcccttcggtattct 30

## RESULT 4

US-08-480-784-75  
; Sequence 75, Application US/08480784  
; Patent No. 5693473

; GENERAL INFORMATION:  
; APPLICANT: Skolnick, Mark H.  
; APPLICANT: Goldgar, David E.  
; APPLICANT: Miki, Yoshio  
; APPLICANT: Swenson, Jeff  
; APPLICANT: Ramb, Alexander  
; APPLICANT: Harshman, Keith D.  
; APPLICANT: Shattuck-Eidens, Donna M.  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Wiseman, Roger W.  
; APPLICANT: Futreal, P. Andrew  
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
; TITLE OF INVENTION: Susceptibility Gene  
; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, N.W., Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,784  
; FILING DATE:

; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,305  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,824  
; FILING DATE: 29-NOV-1994

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/308,104  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/300,266  
; FILING DATE: 02-SEP-1994

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,221  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.

; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109347  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300

; INFORMATION FOR SEQ ID NO: 75:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
US-08-480-784-75

Query Match 63.0%; Score 12.6; DB 1; Length 30;  
Best Local Similarity 78.9%; Pred. No. 8.5e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tctcccttcggtattctt 20

Db 8 TCTCTCTCTCTCTCTCTT 26

RESULT 5  
US-08-483-553-75  
; Sequence 75, Application US/08483553  
; Patent No. 5709999  
; GENERAL INFORMATION:  
; APPLICANT: Skolnick, Mark H.  
; APPLICANT: Goldgar, David E.  
; APPLICANT: Miki, Yoshio  
; APPLICANT: Swenson, Jeff  
; APPLICANT: Kamb, Alexander  
; APPLICANT: Harshman, Keith D.  
; APPLICANT: Shattuck-Eidens, Donna M.  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Wiseman, Roger W.  
; APPLICANT: Futreal, P. Andrew  
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
; TITLE OF INVENTION: Susceptibility Gene  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, N.W., Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,553  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,305  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,824  
; FILING DATE: 29-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/308,104  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/300,266  
; FILING DATE: 02-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,221  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109347  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-08-483-553-75

Query Match 63.08; Score 12.6; DB 1; Length 30;  
Best Local Similarity 78.9%; Pred. No. 8.5e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 tctcccttcggtattatt 20  
||||| ||||| | |||||  
Db 8 TCTCTCTTCCTCTCTT 26  
RESULT 6  
US-08-487-002-75  
; Sequence 75, Application US/08487002  
; Patent No. 5710001  
; GENERAL INFORMATION:  
; APPLICANT: Shattuck-Eidens, Donna M.  
; APPLICANT: Simard, Jacques  
; APPLICANT: Emi, Mitsuru  
; APPLICANT: Nakamura, Yusuke  
; APPLICANT: Durocher, Francine  
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
; TITLE OF INVENTION: Susceptibility Gene  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, N.W., Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,002  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,305  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,824  
; FILING DATE: 29-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/308,104  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/300,266  
; FILING DATE: 02-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,221  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109347  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-08-487-002-75

Query Match 63.08; Score 12.6; DB 1; Length 30;  
Best Local Similarity 78.9%; Pred. No. 8.5e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 tctccctccgttattctt 20  
||||| ||||| | |||||  
Db 8 TCTCTCTCTCTCTCTCTCT 26

RESULT 7  
US-08-483-554B-75  
; Sequence 75, Application US/08483554B  
; Patent No. 5747282  
; GENERAL INFORMATION:  
; APPLICANT: Skolnick, Mark H.  
; APPLICANT: Goldgar, David E.  
; APPLICANT: Miki, Yoshio  
; APPLICANT: Swenson, Jeff  
; APPLICANT: Kamb, Alexander  
; APPLICANT: Harshman, Keith D.  
; APPLICANT: Shattuck-Eidens, Donna M.  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Wiseman, Roger W.  
; APPLICANT: Futreal, P. Andrew  
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
; TITLE OF INVENTION: Susceptibility Gene  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, N.W., Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,554B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,305  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,824  
; FILING DATE: 29-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/308,104  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/300,266  
; FILING DATE: 02-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,221  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109347  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens

Query Match 63.0%; Score 12.6; DB 1; Length 30;  
Best Local Similarity 78.9%; Pred No. 8.5e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 tctccctccgttattctt 20  
||||| ||||| | |||||  
Db 8 TCTCTCTCTCTCTCTCTCT 26

RESULT 8  
US-08-488-011B-75  
; Sequence 75, Application US/08488011B  
; Patent No. 5753441  
; GENERAL INFORMATION:  
; APPLICANT: Skolnick, Mark H.  
; APPLICANT: Goldgar, David E.  
; APPLICANT: Miki, Yoshio  
; APPLICANT: Swenson, Jeff  
; APPLICANT: Kamb, Alexander  
; APPLICANT: Harshman, Keith D.  
; APPLICANT: Shattuck-Eidens, Donna M.  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Wiseman, Roger W.  
; APPLICANT: Futreal, P. Andrew  
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
; TITLE OF INVENTION: Susceptibility Gene  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, N.W., Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,011B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,305  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,824  
; FILING DATE: 29-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/308,104  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/300,266  
; FILING DATE: 02-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,221  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109347-09  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
US-08-488-011B-75

Query Match 63.0%; Score 12.6; DB 1; Length 30;  
Best Local Similarity 78.9%; Pred. No. 8.5e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tctcccttcggtattctt 20  
||||| ||||| |||||  
Db 8 TCTCTCTTCTCTCTCTT 26

RESULT 9  
US-08-850-727-75  
; Sequence 75, Application US/08850727  
; Patent No. 6162897

; GENERAL INFORMATION:  
; APPLICANT: Skolnick, Mark H.  
; APPLICANT: Goldgar, David E.  
; APPLICANT: Miki, Yoshio  
; APPLICANT: Swenson, Jeff  
; APPLICANT: Kamb, Alexander  
; APPLICANT: Harshman, Keith D.  
; APPLICANT: Shattuck-Eidens, Donna M.  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Wiseman, Roger W.  
; APPLICANT: Futreal, P. Andrew  
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, N.W., Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/850,727  
; FILING DATE:

; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/483,554  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,824  
; FILING DATE: 29-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/308,104  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/300,266  
; FILING DATE: 02-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,221  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109347  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-8300  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 75:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
US-08-850-727-75

Query Match 63.0%; Score 12.6; DB 4; Length 30;  
Best Local Similarity 78.9%; Pred. No. 8.5e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tctcccttcggtattctt 20  
||||| ||||| |||||  
Db 8 TCTCTCTTCTCTCTCTT 26

RESULT 10  
PCT-US95-10202-75  
; Sequence 75, Application PC/YUS9510202  
; GENERAL INFORMATION:  
; APPLICANT: Shattuck-Eidens, Donna M.  
; APPLICANT: Simard, Jacques  
; APPLICANT: Emi, Mitsuru  
; APPLICANT: Nakamura, Yusuke  
; APPLICANT: Durocher, Francine  
; TITLE OF INVENTION: In Vivo Mutations and Polymorphisms  
; TITLE OF INVENTION: In the 17q-Linked Breast and Ovarian Cancer  
; TITLE OF INVENTION: Susceptibility Gene  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, N.W., Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/10202  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,305  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,824  
; FILING DATE: 29-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08-308,104  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/300,266  
; FILING DATE: 02-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,221  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109347  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 75:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
PCT-US95-10202-75

Query Match 63.0%; Score 12.6; DB 5; Length 30;  
Best Local Similarity 78.9%; Pred. No. 8.5e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 tctcccttcggtattctt 20  
||||| ||||| | |||||  
Db 8 TCTCTCTCTCTCTCTT 26

RESULT 11  
PCT-US95-10203-75  
Sequence 75, Application PC/TUS9510203  
GENERAL INFORMATION:  
APPLICANT: Skolnick, Mark H.  
APPLICANT: Goldgar, David E.  
APPLICANT: Miki, Yoshio  
APPLICANT: Swenson, Jeff  
APPLICANT: Kamb, Alexander  
APPLICANT: Harshman, Keith D.  
APPLICANT: Shattuck-Eidens, Donna M.  
APPLICANT: Tavtigian, Sean V.  
APPLICANT: Wiseman, Roger W.  
APPLICANT: Futreal, P. Andrew  
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
TITLE OF INVENTION: Susceptibility Gene  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, N.W., Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10203  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,305  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/308,104  
FILING DATE: 16-SEP-1994  
APPLICATION NUMBER: US 08/300,266  
FILING DATE: 02-SEP-1994  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/289,221  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109347  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 75:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
PCT-US95-10203-75

Query Match 63.0%; Score 12.6; DB 5; Length 30;  
Best Local Similarity 78.9%; Pred. No. 8.5e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 tctcccttcggtattctt 20  
||||| ||||| | |||||  
Db 8 TCTCTCTCTCTCTCTT 26

RESULT 12  
PCT-US95-10220-75  
Sequence 75, Application PC/TUS9510220  
GENERAL INFORMATION:  
APPLICANT: Skolnick, Mark H.  
APPLICANT: Goldgar, David E.  
APPLICANT: Miki, Yoshio  
APPLICANT: Swenson, Jeff  
APPLICANT: Kamb, Alexander  
APPLICANT: Harshman, Keith D.  
APPLICANT: Shattuck-Eidens, Donna M.  
APPLICANT: Tavtigian, Sean V.  
APPLICANT: Wiseman, Roger W.  
APPLICANT: Futreal, P. Andrew  
TITLE OF INVENTION: Method for Diagnosing a  
TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, N.W., Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10220  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,305  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/308,104  
FILING DATE: 16-SEP-1994  
APPLICATION NUMBER: US 08/300,266  
FILING DATE: 02-SEP-1994  
PRIOR APPLICATION DATA:

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08-308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Innen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; PCT-US95-10220-75

Query Match 63.0%; Score 12.6; DB 5; Length 30;
Best Local Similarity 78.9%; Pred. No. 8.5e+02;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 tctcccttcggtattctt 20
   ||||| ||||| |||||
DB 8 TCTCTCTTCCTCTCTT 26

RESULT 13
US-08-701-380-4
; Sequence 4, Application US/08701380
; Patent No. 5686598
; GENERAL INFORMATION:
; APPLICANT: NORTH, Michael
; APPLICANT: NISHINA, Patsy
; APPLICANT: NAGERT, Juergen
; TITLE OF INVENTION: GENES ASSOCIATED WITH RETINAL
; TITLE OF INVENTION: DYSTROPHIES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/701,380
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: A-63565/PJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-8700
; TELEFAX: 415-494-8771
; TELEX: 910 277299
```

```
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primers"
US-08-701-380-4

Query Match 61.0%; Score 12.2; DB 1; Length 22;
Best Local Similarity 82.4%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tctcccttcggtatttc 18
   ||||| ||||| |||||
DB 2 TCTCCCTTCCTCTCTTC 18

RESULT 14
US-09-032-365A-38
; Sequence 38, Application US/09032365A
; Patent No. 6114502
; GENERAL INFORMATION:
; APPLICANT: No. 6114502th, Michael
; APPLICANT: Nishina, Patsy
; APPLICANT: Naggart, Juergen
; APPLICANT: No. 6114502en-Trauth, Konrad
; TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
; TITLE OF INVENTION: NEUROSENSORY DEFECTS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,365A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-2CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650 327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-032-365A-38

Query Match 61.0%; Score 12.2; DB 3; Length 22;
Best Local Similarity 82.4%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Qy 2 tctcccttcggtattc 18  
||||| | | |  
Db 2 TCTCCCTTCCTTCCTTC 18

## RESULT 15

US-07-931-473B-35/c  
; Sequence 35, Application US/07931473B  
; Patent No. 5270163  
; GENERAL INFORMATION:  
; APPLICANT: Larry Gold  
; APPLICANT: Craig Tuerk  
; TITLE OF INVENTION: Nucleic Acid Ligands  
; NUMBER OF SEQUENCES: 335  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Beaton & Swanson, P.C.  
; STREET: 4582 South Ulster Street Parkway, #403  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80237  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07931.473B  
; FILING DATE: 19920817  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barry J. Swanson  
; REGISTRATION NUMBER: 33,215  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 850-9900  
; TELEFAX: (303) 850-9401  
; INFORMATION FOR SEQ ID NO: 35:  
; . SEQUENCE CHARACTERISTICS:  
; . LENGTH: 30 nucleotides  
; . TYPE: NUCLEIC ACID  
; . STRANDEDNESS: single  
; . TOPOLOGY: linear  
US-07-931-473B-35

Query Match 60.0%; Score 12; DB 1; Length 30;  
Best Local Similarity 75.0%; Pred. No. 1.6e+03;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tctcccttcggtattc 20  
||||| | | |  
Db 25 TTCCTTCCTTCCTTCCTTC 6

Search completed: October 2, 2001, 16:03:47  
Job time: 14591 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:41 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-8

Perfect score: 20

Sequence: 1 ctgatatgctaggtatctgtc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_ba3:\*
- 4: gb\_in1:\*
- 5: gb\_in2:\*
- 6: gb\_in3:\*
- 7: gb\_om:\*
- 8: gb\_ov:\*
- 9: gb\_pat1:\*
- 10: gb\_pat2:\*
- 11: gb\_ph:\*
- 12: gb\_pil:\*
- 13: gb\_pl2:\*
- 14: gb\_pl3:\*
- 15: gb\_pl4:\*
- 16: em\_ba1:\*
- 17: em\_ba2:\*
- 18: em\_fun:\*
- 19: em\_htgo\_hum:\*
- 20: em\_htgo\_inv:\*
- 21: em\_htgo\_rod:\*
- 22: em\_htg\_hum1:\*
- 23: em\_htg\_hum2:\*
- 24: em\_htg\_hum3:\*
- 25: em\_htg\_hum4:\*
- 26: em\_htg\_hum5:\*
- 27: em\_htg\_hum6:\*
- 28: em\_htg\_hum7:\*
- 29: em\_htg\_hum8:\*
- 30: em\_htg\_inv1:\*
- 31: em\_htg\_inv2:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_rod:\*
- 34: em\_hum1:\*
- 35: em\_hum2:\*
- 36: em\_hum3:\*
- 37: em\_hum4:\*
- 38: em\_hum5:\*
- 39: em\_hum6:\*
- 40: em\_hum7:\*
- 41: em\_in:\*
- 42: em\_om:\*
- 43: em\_or:\*

- 44: em\_ov:\*
- 45: em\_pat:\*
- 46: em\_ph:\*
- 47: em\_pl:\*
- 48: em\_ro:\*
- 49: em\_sts:\*
- 50: em\_sy:\*
- 51: em\_un:\*
- 52: em\_vi:\*
- 53: gb\_sts1:\*
- 54: gb\_sts2:\*
- 55: gb\_sts3:\*
- 56: gb\_sy:\*
- 57: gb\_un:\*
- 58: gb\_vil:\*
- 59: gb\_vil2:\*
- 60: gb\_vil2:\*
- 61: gb\_vil2:\*
- 62: gb\_vil2:\*
- 63: gb\_vil2:\*
- 64: gb\_vil2:\*
- 65: gb\_vil2:\*
- 66: gb\_vil2:\*
- 67: gb\_vil2:\*
- 68: gb\_vil2:\*
- 69: gb\_vil2:\*
- 70: gb\_vil2:\*
- 71: gb\_vil2:\*
- 72: gb\_vil2:\*
- 73: gb\_vil2:\*
- 74: gb\_vil2:\*
- 75: gb\_vil2:\*
- 76: gb\_vil2:\*
- 77: gb\_vil2:\*
- 78: gb\_vil2:\*
- 79: gb\_vil2:\*
- 80: gb\_vil2:\*
- 81: gb\_vil2:\*
- 82: gb\_vil2:\*
- 83: gb\_vil2:\*
- 84: gb\_vil2:\*
- 85: gb\_vil2:\*
- 86: gb\_vil2:\*
- 87: gb\_vil2:\*
- 88: gb\_vil2:\*
- 89: gb\_vil2:\*
- 90: gb\_vil2:\*
- 91: gb\_vil2:\*
- 92: gb\_vil2:\*
- 93: gb\_vil2:\*
- 94: gb\_vil2:\*
- 95: gb\_vil2:\*
- 96: gb\_vil2:\*
- 97: gb\_vil2:\*
- 98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
c 1	13.4	67.0	43	9 AR079693	Sequence AR079693
c 2	13.4	67.0	43	9 AR081223	Sequence AR081223
c 3	12.8	64.0	24	9 AG3162	Sequence AG3162
4	12.6	63.0	21	9 AR001344	Sequence AR001344
5	12.6	63.0	21	9 AR037524	Sequence AR037524
6	12.6	63.0	21	9 AR062804	Sequence AR062804
7	12.6	63.0	21	9 AR087882	Sequence AR087882
8	12.6	63.0	21	9 AR091352	Sequence AR091352

```

9 12.6 63.0 21 10 I41178 Sequence 34
10 12.6 63.0 27 9 AR012082 Sequence
11 12.6 63.0 27 9 AR012083 Sequence
12 12.4 62.0 35 9 AR030753 Sequence
13 12.4 62.0 37 10 AX100491 Sequence
14 12.2 61.0 42 9 AR054989 Sequence
15 12.2 61.0 44 9 AR036477 Sequence
16 12.2 61.0 44 9 AR069436 Sequence
17 12.2 61.0 44 10 I73458 Sequence 10
18 12.2 60.0 25 9 AX060728 Sequence
19 12.2 60.0 25 9 AX060907 Sequence
20 12.2 60.0 36 10 I09627 Sequence 1
21 12.2 60.0 45 93 HSTRK3X22
22 11.8 59.0 22 9 AR053115
23 11.8 59.0 39 9 A36972 Sequence 23
24 11.8 59.0 39 9 A39180 Sequence 28
25 11.8 59.0 39 9 A39239 Sequence 23
26 11.8 59.0 40 9 A46304 Sequence 2
27 11.8 59.0 40 9 A80185 Sequence 2
28 11.8 59.0 45 94 MMU232757
29 11.6 58.0 20 9 AR072271 Sequence
30 11.6 58.0 20 9 AR082328 Sequence
31 11.6 58.0 20 9 AR082345 Sequence
32 11.6 58.0 20 10 I26382 Sequence 74
33 11.6 58.0 20 10 I78374 Sequence 17
34 11.6 58.0 20 10 I78391 Sequence 18
35 11.6 58.0 33 93 HSTCELL11
36 11.6 58.0 35 10 AX100369 Sequence
37 11.6 58.0 36 9 A71632 Sequence 65
38 11.6 58.0 36 10 E16846 PCR primer
39 11.6 58.0 40 9 AR063784 Sequence
40 11.6 58.0 40 10 I30147 Sequence 69
41 11.6 58.0 42 93 HSCD85715
42 11.4 57.0 28 9 A98431 Sequence 15
43 11.2 56.0 20 9 AX008785 Sequence
44 11.2 56.0 23 9 AR020665 Sequence
45 11.2 56.0 24 9 A41626 Sequence 35

```

## ALIGNMENTS

```

RESULT 1
AR079693/c
LOCUS AR079693 43 bp DNA
DEFINITION Sequence 17 from patent US 5965726.
ACCESSION AR079693
VERSION AR079693.1 GI:10006434
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 43)
AUTHORS Pavlakis,G.N. and Felber,B.K.
TITLE Method of eliminating inhibitory/ instability regions of mRNA
JOURNAL Patent: US 5965726-A 17 12-OCT-1999;
FEATURES
Location/Qualifiers
Source
BASE COUNT 9 a 11 c 5 g 18 t
ORIGIN
Query Match 67.0%; Score 13.4; DB 9; Length 43;
Best Local Similarity 93.3%; Pred. No. 4.4e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 gatgctaggtatctg 18
||| ||||| |||||
Db 19 GAAGCTAGGTATCTG 5

RESULT 2
AR079693/c
LOCUS AR079693 43 bp DNA
DEFINITION Sequence 17 from patent US 5972596.
ACCESSION AR079693
VERSION AR079693.1 GI:10007949
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 43)
AUTHORS Pavlakis,G.N. and Felber,B.K.
TITLE Nucleic acid constructs containing HIV genes with mutated inhibitory/instability regions and methods of using same
JOURNAL Patent: US 5972596-A 17 26-OCT-1999;
FEATURES
Location/Qualifiers
Source
BASE COUNT 9 a 11 c 5 g 18 t
ORIGIN
Query Match 67.0%; Score 13.4; DB 9; Length 43;
Best Local Similarity 93.3%; Pred. No. 4.4e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 gatgctaggtatctg 18
||| ||||| |||||
Db 19 GAAGCTAGGTATCTG 5

RESULT 3
A63162
LOCUS A63162 24 bp DNA
DEFINITION Sequence 5 from Patent WO9718307.
ACCESSION A63162
VERSION A63162.1 GI:3717015
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Buehler,T.
TITLE FAS LIGAND FUSION PROTEIN
JOURNAL Patent: WO 9718307-A 5 22-MAY-1997;
COMMENT SANDOZ LTD (CH)
Other publication AU 7684896 19970605.
FEATURES
Location/Qualifiers
Source
BASE COUNT 3 a 5 c 9 g 7 t
ORIGIN
Query Match 64.0%; Score 12.8; DB 9; Length 24;
Best Local Similarity 87.5%; Pred. No. 9.7e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agatgctaggtatctg 18
||||| ||| |||||
Db 8 AGATGCTGGGGATCTG 23

RESULT 4
AR001344
LOCUS AR001344 21 bp DNA
DEFINITION Sequence 34 from patent US 5739027.
ACCESSION AR001344
VERSION AR001344.1 GI:3963411
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 21)
AUTHORS Buehler,T.
TITLE FAS LIGAND FUSION PROTEIN
JOURNAL Patent: WO 9718307-A 5 22-MAY-1997;
COMMENT SANDOZ LTD (CH)
Other publication AU 7684896 19970605.
FEATURES
Location/Qualifiers
Source
BASE COUNT 3 a 5 c 9 g 7 t
ORIGIN
Query Match 64.0%; Score 12.8; DB 9; Length 24;
Best Local Similarity 87.5%; Pred. No. 9.7e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agatgctaggtatctg 18
||||| ||| |||||
Db 8 AGATGCTGGGGATCTG 23

```

REFERENCE 1 (bases 1 to 21)  
AUTHORS Kamb,A.  
TITLE MTS1 $\beta$ .beta. gene  
JOURNAL Patent: US 5739027-A 34 14-APR-1998;  
FEATURES Location/Qualifiers  
source  
1..21  
BASE COUNT 5 a 3 c 7 g 6 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 21;  
Best Local Similarity 78.9%; Pred. No. 1.3e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctatagctcctaggtagctatctgt 19  
||||| ||| | |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

RESULT 5  
LOCUS AR037524 21 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 34 from patent US 5801236.  
ACCESSION AR037524  
VERSION AR037524.1 GI:5955380  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Kamb,A.  
TITLE Probes for MTS1 gene and polynucleotides encoding mutant MTS1 genes  
JOURNAL Patent: US 5801236-A 34 01-SEP-1998;  
FEATURES Location/Qualifiers  
source  
1..21  
BASE COUNT 5 a 3 c 7 g 6 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 21;  
Best Local Similarity 78.9%; Pred. No. 1.3e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctatagctcctaggtagctatctgt 19  
||||| ||| | |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

RESULT 6  
LOCUS AR062804 21 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 34 from patent US 5843756.  
ACCESSION AR062804  
VERSION AR062804.1 GI:5990495  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Stone,S., Jiang,P. and Kamb,A.  
TITLE Mouse MTS1 gene  
JOURNAL Patent: US 5843756-A 34 01-DEC-1998;  
FEATURES Location/Qualifiers  
source  
1..21  
BASE COUNT 5 a 3 c 7 g 6 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 21;  
Best Local Similarity 78.9%; Pred. No. 1.3e+04;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ctatagctcctaggtagctatctgt 19  
||||| ||| | |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

RESULT 7  
LOCUS AR087882 21 bp DNA PAT 07-SEP-2000  
DEFINITION Sequence 34 from patent US 5989815.  
ACCESSION AR087882  
VERSION AR087882.1 GI:10014645  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Skolnick,M.H., Cannon-Albright,L.A. and Kamb,A.  
TITLE Methods for detecting predispotion to cancer at the MTS gene  
JOURNAL Patent: US 5989815-A 34 23-NOV-1999;  
FEATURES Location/Qualifiers  
source  
1..21  
BASE COUNT 5 a 3 c 7 g 6 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 21;  
Best Local Similarity 78.9%; Pred. No. 1.3e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctatagctcctaggtagctatctgt 19  
||||| ||| | |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

RESULT 8  
LOCUS AR091352 21 bp DNA PAT 07-SEP-2000  
DEFINITION Sequence 34 from patent US 5994095.  
ACCESSION AR091352  
VERSION AR091352.1 GI:10018107  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Kamb,A.  
TITLE MTS2 gene  
JOURNAL Patent: US 5994095-A 34 30-NOV-1999;  
FEATURES Location/Qualifiers  
source  
1..21  
BASE COUNT 5 a 3 c 7 g 6 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 21;  
Best Local Similarity 78.9%; Pred. No. 1.3e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctatagctcctaggtagctatctgt 19  
||||| ||| | |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

RESULT 9  
LOCUS I41178 21 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 34 from patent US 5624819.  
ACCESSION I41178  
VERSION I41178.1 GI:2081768

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Skolnick,M.H., Cannon-Albright,L.A. and Kamb,A.  
TITLE Germ-line mutations in the MTS gene  
JOURNAL Patent: US 5624819-A 34 29-APR-1997;  
FEATURES Location/Qualifiers  
1..21  
/organism="unknown"  
BASE COUNT 5 a 3 c 7 g 6 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 21;  
Best Local Similarity 78.9%; Pred. No. 1.3e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 19  
||||| ||| |||||  
Db 3 CTAGAGCGCAATTACTGT 21

RESULT 10  
AR012082 27 bp DNA PAT 04-DEC-1998  
LOCUS AR012082  
DEFINITION Sequence 6 from patent US 5763190.  
ACCESSION AR012082  
VERSION AR012082.1 GI:3970072  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 27)  
AUTHORS Weiner,D.B. and Refaelli,Y.  
TITLE Methods for the identification of compounds capable of inducing the nuclear translocation of a receptor complex comprising the glucocorticoid receptor type II and viral protein R interacting protein  
JOURNAL Patent: US 5763190-A 6 09-JUN-1998;  
FEATURES Location/Qualifiers  
1..27  
/organism="unknown"  
BASE COUNT 6 a 6 c 8 g 7 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 27;  
Best Local Similarity 78.9%; Pred. No. 1.3e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagatgctaggtatctgtc 20  
||||| ||||| |||||  
Db 8 TAGTTGCTAGGCAACTGGC 26

RESULT 11  
AR012083/c 27 bp DNA PAT 04-DEC-1998  
LOCUS AR012083  
DEFINITION Sequence 7 from patent US 5763190.  
ACCESSION AR012083  
VERSION AR012083.1 GI:3970073  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 27)  
AUTHORS Weiner,D.B. and Refaelli,Y.  
TITLE Methods for the identification of compounds capable of inducing the nuclear translocation of a receptor complex comprising the glucocorticoid receptor type II and viral protein R interacting protein

JOURNAL Patent: US 5763190-A 7 09-JUN-1998;  
FEATURES Location/Qualifiers  
1..27  
/organism="unknown"  
BASE COUNT 7 a 8 c 6 g 6 t  
ORIGIN

Query Match 62.0%; Score 12.4; DB 9; Length 35;  
Best Local Similarity 65.0%; Pred. No. 1.7e+04;  
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20  
||||| ||||| |||||  
Db 16 CTGCGCGCAGGCTCTGYGTC 35

RESULT 13  
AX100491/c 37 bp DNA PAT 10-APR-2001  
LOCUS AX100491  
DEFINITION Sequence 5 from Patent WO0121783.  
ACCESSION AX100491  
VERSION AX100491.1 GI:13619505  
KEYWORDS  
SOURCE zea mays.  
ORGANISM zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 37)  
AUTHORS Jiao,S., Habben,J.E. and Niu,X.  
TITLE Seed-preferred promoter from maize  
JOURNAL Patent: WO 0121783-A 5 29-MAR-2001;  
PIONEER HI-BRED INTERNATIONAL, INC. (US)  
FEATURES Location/Qualifiers  
1..37  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
BASE COUNT 9 a 10 c 7 g 11 t  
ORIGIN

JOURNAL Patent: US 5763190-A 7 09-JUN-1998;  
FEATURES Location/Qualifiers  
1..27  
/organism="unknown"  
BASE COUNT 7 a 8 c 6 g 6 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 27;  
Best Local Similarity 78.9%; Pred. No. 1.3e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagatgctaggtatctgtc 20  
||||| ||||| |||||  
Db 24 TAGTTGCTAGGCAACTGGC 6

RESULT 12  
AR030753 35 bp DNA PAT 29-SEP-1999  
LOCUS AR030753  
DEFINITION Sequence 2 from patent US 5861301.  
ACCESSION AR030753  
VERSION AR030753.1 GI:5943967  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 35)  
AUTHORS Terman,B.,Israel and Carrion,M.Eduardo.  
TITLE Recombinant kinase insert domain containing receptor and gene encoding same  
JOURNAL Patent: US 5861301-A 2 19-JAN-1999;  
FEATURES Location/Qualifiers  
1..35  
/organism="unknown"  
BASE COUNT 5 a 9 c 8 g 5 others  
ORIGIN

Query Match 62.0%; Score 12.4; DB 9; Length 35;  
Best Local Similarity 65.0%; Pred. No. 1.7e+04;  
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20  
||||| ||||| |||||  
Db 16 CTGCGCGCAGGCTCTGYGTC 35

RESULT 13  
AX100491/c 37 bp DNA PAT 10-APR-2001  
LOCUS AX100491  
DEFINITION Sequence 5 from Patent WO0121783.  
ACCESSION AX100491  
VERSION AX100491.1 GI:13619505  
KEYWORDS  
SOURCE zea mays.  
ORGANISM zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 37)  
AUTHORS Jiao,S., Habben,J.E. and Niu,X.  
TITLE Seed-preferred promoter from maize  
JOURNAL Patent: WO 0121783-A 5 29-MAR-2001;  
PIONEER HI-BRED INTERNATIONAL, INC. (US)  
FEATURES Location/Qualifiers  
1..37  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
BASE COUNT 9 a 10 c 7 g 11 t  
ORIGIN

Search completed: October 2, 2001, 15:56:42  
Job time: 14166 sec

Query Match 62.0%; Score 12.4; DB 10; Length 37;  
Best Local Similarity 92.9%; Pred. No. 1.7e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ctagatgctaggta 14  
||||| |||||  
Db 37 CTAGATGATAGGTA 24

## RESULT 14

AR054989 AR054989 42 bp DNA PAT 29-SEP-1999  
LOCUS  
DEFINITION Sequence 2 from patent US 5837464.  
ACCESSION AR054989  
VERSION AR054989.1 GI:5980566  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 42)  
AUTHORS Capon,D. and Petropoulos,C.J.  
TITLE Compositions and methods for determining anti-viral drug  
susceptibility and resistance and anti-viral drug screening  
JOURNAL Patent: US 5837464-A 2 17-NOV-1998;  
FEATURES Location/Qualifiers  
source 1..42  
BASE COUNT 10 a 8 c 12 g 12 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 42;  
Best Local Similarity 82.4%; Pred. No. 2.2e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ctagatgctagggtatct 17  
||||| ||||| |||  
Db 20 CTAGATGCTAGAGATTT 36

## RESULT 15

AR036477/c AR036477 44 bp DNA PAT 29-SEP-1999  
LOCUS  
DEFINITION Sequence 10 from patent US 5872218.  
ACCESSION AR036477  
VERSION AR036477.1 GI:5953145  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 44)  
AUTHORS Wolf,D., Tomlinson,J.E., Fretto,L.J., Giese,N.A., Escobedo,J.A. and  
Williams,L.Thomas.  
TITLE Human platelet-derived growth factor receptor extracellular domain  
antibodies  
JOURNAL Patent: US 5872218-A 10 16-FEB-1999;  
FEATURES Location/Qualifiers  
source 1..44  
BASE COUNT 6 a 13 c 11 g 14 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 44;  
Best Local Similarity 82.4%; Pred. No. 2.3e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gatgctagggtatctgtc 20  
||||| |||||  
Db 31 GAGCTAGGGATCTGGC 15

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:00:51 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-8

Perfect score: 20

Sequence: 1 ctatagctagtagtctgtc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est75:\*  
137: gb\_est76:\*  
138: gb\_est77:\*  
139: gb\_est78:\*  
140: gb\_est79:\*  
141: gb\_est80:\*  
142: gb\_est81:\*  
143: gb\_est82:\*  
144: gb\_est83:\*  
145: gb\_est84:\*  
146: gb\_est85:\*  
147: gb\_est86:\*  
148: gb\_est87:\*  
149: gb\_est88:\*  
150: gb\_est89:\*  
151: gb\_est90:\*  
152: gb\_est99:\*  
153: gb\_est100:\*  
154: gb\_est101:\*  
155: gb\_est102:\*  
156: gb\_est103:\*  
157: gb\_est104:\*  
158: gb\_est105:\*  
159: gb\_est106:\*  
160: gb\_est67:\*  
161: gb\_est68:\*  
162: gb\_est69:\*  
163: gb\_est70:\*  
164: gb\_est71:\*  
165: gb\_est72:\*  
166: gb\_est73:\*  
167: gb\_est74:\*  
168: gb\_est91:\*  
169: gb\_est92:\*  
170: gb\_est93:\*  
171: gb\_est94:\*  
172: gb\_est95:\*  
173: gb\_est96:\*  
174: gb\_est97:\*  
175: gb\_est98:\*  
176: em\_esthum29:\*  
177: em\_esthum30:\*  
178: em\_esthum31:\*  
179: em\_esthum32:\*  
180: em\_esthum33:\*  
181: em\_estom3:\*  
182: em\_estp11:\*  
183: em\_estro21:\*  
184: em\_estro22:\*  
185: em\_estro23:\*  
186: em\_htc:\*  
187: gb\_est107:\*  
188: gb\_est108:\*  
189: gb\_est109:\*

190: gb\_est110:\*  
191: gb\_est111:\*  
192: gb\_htc:\*  
193: em\_gss\_fun:\*  
194: em\_gss\_hum1:\*  
195: em\_gss\_hum2:\*  
196: em\_gss\_hum3:\*  
197: em\_gss\_hum4:\*  
198: em\_gss\_hum5:\*  
199: em\_gss\_hum6:\*  
200: em\_gss\_hum7:\*  
201: em\_gss\_hum8:\*  
202: em\_gss\_hum9:\*  
203: em\_gss\_inv1:\*  
204: em\_gss\_inv2:\*  
205: em\_gss\_inv3:\*  
206: em\_gss\_other:\*  
207: em\_gss\_pln1:\*  
208: em\_gss\_pln2:\*  
209: em\_gss\_pro:\*  
210: em\_gss\_rod1:\*  
211: em\_gss\_rod2:\*  
212: em\_gss\_rod3:\*  
213: em\_gss\_rod4:\*  
214: em\_gss\_rod5:\*  
215: em\_gss\_vrt1:\*  
216: em\_gss\_vrt2:\*  
217: em\_gss\_vrt3:\*  
218: gb\_gss1:\*  
219: gb\_gss2:\*  
220: gb\_gss3:\*  
221: gb\_gss4:\*  
222: gb\_gss5:\*  
223: gb\_gss6:\*  
224: gb\_gss7:\*  
225: gb\_gss8:\*  
226: gb\_gss9:\*  
227: gb\_gss10:\*  
228: gb\_gss11:\*  
229: gb\_gss12:\*  
230: gb\_gss13:\*  
231: gb\_gss14:\*  
232: gb\_gss15:\*  
233: gb\_gss16:\*  
234: gb\_gss17:\*  
235: gb\_gss18:\*  
236: gb\_gss19:\*  
237: gb\_gss20:\*  
238: gb\_gss21:\*  
239: gb\_gss22:\*  
240: gb\_gss23:\*  
241: gb\_gss24:\*  
242: gb\_gss25:\*  
243: gb\_gss26:\*  
244: gb\_gss27:\*  
245: gb\_gss28:\*  
246: gb\_gss29:\*  
247: gb\_gss30:\*  
248: gb\_gss31:\*  
249: gb\_gss32:\*  
250: gb\_gss33:\*  
251: gb\_gss34:\*  
252: em\_gss\_inv4:\*  
253: em\_gss\_rod6:\*  
254: em\_gss\_rod7:\*  
255: em\_gss\_rod8:\*  
256: gb\_gss35:\*  
257: gb\_gss36:\*  
258: gb\_gss37:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



SUMMARIES				TITLE	
Result No.	Score	Query Match	Length DB ID	JOURNAL COMMENT	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
1	12.8	64.0	48 251	Unpublished (2000)	plasmid inserts
2	12.2	61.0	37 21	Contact: Robert B. Weiss	Unpublished (2000)
3	12	60.0	49 12	University of Utah Genome Center	University of Utah
4	11.8	59.0	44 141	Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA	
5	11.8	59.0	49 189	Tel: 801 585 5606	Fax: 801 585 7177
6	11.8	59.0	50 243	Email: dunn@genetics.utah.edu	
7	11.6	58.0	40 21	Insert Length: 10000 Std Error: 0.00	
8	11.6	58.0	41 244	Plate: 0180 row: L column: 11	
9	11.6	58.0	45 250	Seq primer: CGTTGTAACGACGCGCAGT	
10	11.6	58.0	50 107	Class: plasmid ends	
11	11.6	58.0	50 107	High quality sequence stop: 48.	
12	11.6	58.0	50 107	Location/Qualifiers	
13	11.6	58.0	50 107	1. 48	
14	11.4	57.0	36 4	/organism="Mus musculus"	
15	11.4	57.0	39 243	/strain="C57BL/6J"	
16	11.4	57.0	40 247	/db_xref="taxon:10090"	
17	11.2	56.0	33 243	/clone="UUGC2M0180L11"	
18	11.2	56.0	36 189	/clone_lib="Mouse 10kb plasmid UUGC1M library"	
19	11.2	56.0	37 15	/sex="Male"	
20	11.2	56.0	46 104	/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"	
21	11.2	56.0	49 243	/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource	
22	11.2	56.0	49 246	(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pW042 (gil4732114 gb AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."	
23	11.2	56.0	50 107	BASE COUNT	10 a 10 c 16 g 12 t
24	11.2	56.0	50 107	ORIGIN	
25	11.2	56.0	50 107	Query Match	64.0%; Score 12.8; DB 251; Length 48;
26	11.2	56.0	50 107	Best Local Similarity	87.5%; Pred. No. 1.7e+04;
27	11.2	56.0	50 107	Matches	14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
28	11.5	55.0	31 9	QY	4 gatgctagggtatctgt 19
29	11.5	55.0	36 242	Db	
30	11.5	55.0	38 249	RESULT 2	
31	11.5	55.0	45 244	AI545803	AI545803 37 bp mRNA EST 22-MAR-1999
32	11.5	55.0	47 246	LOCUS	fb65612.y1 zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
33	11.5	55.0	49 2	DEFINITION	TR:Q31365 Q31365 UBA*01 CLASS I MHC MRNA. ;, mRNA sequence.
34	11.5	55.0	50 12	ACCESSION	AI545803
35	11.5	55.0	50 12	VERSION	AI545803.1
36	10.8	54.0	21 246	KEYWORDS	EST. GI:4463176
37	10.8	54.0	31 139	SOURCE	zebrafish.
38	10.8	54.0	31 139	ORGANISM	Danio rerio
39	10.8	54.0	33 143	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
40	10.8	54.0	39 242	AUTHORS	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
41	10.8	54.0	45 106		Cypriniformes; Cyprinidae; Rasbora; Danio.
42	10.8	54.0	45 106		1 (bases 1 to 37)
43	10.8	54.0	46 13		Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
44	10.8	54.0	49 20		, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
45	10.8	54.0	50 107		

K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

# TITLE JOURNAL COMMENT

WashU Zebrafish EST Project 1998  
Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu  
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
Ressourcenzentrum Primardatenbank, Berlin, Germany (web address:  
www.rzpd.de)

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: T3 ET from Amersham

High quality sequence stop: 1.

## FEATURES

### source

1. .37  
/organism="Danio rerio"  
/db.xref="taxon:7955"  
/clone.lib="Zebrafish WashU MPIMG EST"  
/sex="mixed"  
/tissue.type="26 somite embryos, adult livers, shield  
stage embryos"  
/lab.host="XLI-blue MRF"  
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st  
strand cDNA was primed with a Not I - oligo(dT)15 primer  
15'GCACAGTCTTAGTCGCGCGCGCCCTTTTCTTTT3';  
double-stranded cDNA was ligated to Sal I adaptors (BRL),  
digested with Not I and cloned into the Not I and Sal I  
sites of the pSPORT1 vector (BRL). Library was constructed  
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck  
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
analysis were selected following oligonucleotide  
hybridization fingerprinting of arrayed clones from  
zebrafish late somitogenesis (26 ss), adult liver or  
embryonic shield stage (5.6 h) libraries. Fingerprint  
data were used to computationally cluster cDNAs, and a  
single cDNA from each cluster was chosen for sequencing.  
In some cases multiple members of the same cluster were  
sequenced to assess clustering parameters or single clones  
were sequenced additional times to assess quality  
control."

BASE COUNT 5 a 8 c 9 g 15 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 21; Length 37;

Best Local Similarity 82.4%; Pred. NO. 3.3e+04;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 gatgctaggtatctgtc 20

| | | | | | | | | | | | | | | |

Db 19 GCTGCTAGTCTTCTGTC 35

## RESULT 3

AA797125 49 bp mRNA EST 10-FEB-1998  
LOCUS vn3je06.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:1023970 5' similar to SW:TRHY\_RABIT P37709 TRICHOHYALIN. [1]  
; mRNA sequence.  
AA797125  
ACCESSION AA797125.1 GI:2860080  
VERSION  
KEYWORDS EST.

## SOURCE ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 49)  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:574746

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 revl ET from Amersham

High quality sequence stop: 1.

## FEATURES

### source

1. .49  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db.xref="taxon:10090"  
/clone.lib="IMAGE:1023970"  
/clone.lib="Stratagene mouse skin (#937313)"  
/sex="females"  
/tissue.type="whole skin"  
/dev.stage="11 weeks old"  
/lab.host="SOLR (kanamycin resistant)"  
/note="Organ: skin; Vector: pBluescript SK-; Site 1: EcoRI  
; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5',  
adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3"

BASE COUNT 12 a 12 c 14 g 11 t

## ORIGIN

Query Match 60.0%; Score 12; DB 12; Length 49;

Best Local Similarity 75.0%; Pred. No. 4.4e+04;

Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctgatgctaggtatctgtc 20

| | | | | | | | | | | | | | | |

Db 8 CGAGATGCTCAATACCTGTC 27

## RESULT 4

BE900863

LOCUS BE900863

DEFINITION 601673838P1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3956864 5',

mRNA sequence.

ACCESSION BE900863

VERSION BE900863.1 GI:10389463

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 44)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLC833 row: k column: 09.

## FEATURES

source

1. .44  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3956864"  
 /clone\_lib="NIH\_MGC\_21"  
 /tissue\_type="choriocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: placenta; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT  
 ORIGIN

11 a

7 c 16 g 10 t

Query Match 59.0%; Score 11.8; DB 141; Length 44;  
 Best Local Similarity 86.7%; Pred. No. 5.5e+04;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tagatgctaggtatc 16

||| |||||

Db 14 TAGGGGCTAGGTATC 28

RESULT 5

LOCUS

T56680 49 bp mRNA EST 07-FEB-1995  
 ya95b09.r1 Stratagene placenta (#937225) Homo sapiens cDNA clone  
 IMAGE:69401 5' similar to similar to gb:M22995 RAS-RELATED PROTEIN  
 RAP-1A (HUMAN), mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 49)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins  
 M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore  
 B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,  
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,  
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)

TITLE

MEDLINE

COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
 Insert Size: 446

High quality sequence starts: 1 High quality sequence stops: 1  
 Source: IMAGE Consortium, LLNL This clone is available royalty-free  
 through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov))  
 for further information. Trace considered overall poor quality  
 Insert Length: 446 Std Error: 0.00  
 Seq primer: M13RPI  
 High quality sequence stop: 1.

## FEATURES

source

1. .49  
 /organism="Homo sapiens"  
 /db\_xref="GDB:490298"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:69401"  
 /clone\_lib="Stratagene placenta (#937225)"  
 /sex="male"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: placenta; Vector: pBluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. Caucasian. Average insert size: 1.2 kb; Uni-ZAP  
 XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAG 3' ~3'  
 adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT

ORIGIN

12 a

8 c 7 g 15 t

Query Match 59.0%; Score 11.8; DB 189; Length 49;  
 Best Local Similarity 81.2%; Pred. No. 5.6e+04;  
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 atgtcaggtatctgtc 20

||| |||||

Db 24 ATNCTTGTATCTGTC 39

RESULT 6

LOCUS

AZ423820 50 bp DNA GSS 03-OCT-2000  
 IM0203P22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0203P22 F, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus  
 house mouse  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 50)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.

TITLE

JOURNAL

COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0203 row: p column: 22  
 Seq primer: CGTGTAAACGACGGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 50.

## FEATURES

source

1. .50  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0203P22"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: pWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance." 9 t

BASE COUNT 15 a 15 c 11 g 9 t

ORIGIN

Query Match 59.0%; Score 11.8; DB 243; Length 50;  
Best Local Similarity 86.7%; Pred. No. 5.6e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 6 tgcgtaggatctgtc 20  
||| ||||| |||||  
Db 49 TGTAGGTCGTCTGTC 35

## RESULT 7

A1499269 40 bp mRNA EST 11-MAR-1999  
LOCUS to09c08.x1 NCI-CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:2178542 3'  
DEFINITION similar to TR:Q12803 Q12803 LEUKOPHYSIN.; contains element MER40  
repetitive element ;, mRNA sequence.

ACCESSION A1499269  
VERSION A1499269.1 GI:4391251  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 40)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.

## FEATURES

## source

1..40  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2178542"  
/clone\_lib="NCI-CGAP\_Ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"

BASE COUNT 10 a 9 c 11 g 10 t

ORIGIN

Query Match 58.0%; Score 11.6; DB 21; Length 40;  
Best Local Similarity 77.8%; Pred. No. 6.9e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 agatgctaggatctgtc 20  
||| ||||| |||||  
Db 4 AGGGGCTAGGAATCTGAC 21

## RESULT 8

AZ471345/c 41 bp DNA GSS 04-OCT-2000  
LOCUS 1M0285I19R Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
DEFINITION Clone UUGCLM0285I19 R, DNA sequence.

ACCESSION AZ471345  
VERSION AZ471345.1 GI:10629470  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 41)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0285 row: 1 column: 19  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 41.

## FEATURES

## source

1..41  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCLM0285I19"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 11 a 9 c 11 g 10 t

ORIGIN



TITLE H., Ota.T., Isogai.T., Tanaka.T., Nakamura.Y., Morishita.S., Okubo  
K., Suyama.A. and Sugano.S.  
JOURNAL Fine Structural analysis of transcription start sites of human  
COMMENT mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
Unpublished (2001)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp  
Suzuki.Y., Yoshitomo-Nakagawa.K., Maruyama.K., Suyama.A. and Sugano  
S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES  
source  
1..50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CAS09268"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 19 a 11 c 12 g 8 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 107; Length 50;  
Best Local Similarity 77.8%; Pred. No. 7.2e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 agatgctaggtagctgtc 20  
||||| ||||| |||||

Db 43 AGAGGTTGGTATCTGCC 26

RESULT 12  
LOCUS AU102843 50 bp mRNA EST 05-APR-2001  
DEFINITION AU102843 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
CAS11158, mRNA sequence.  
ACCESSION AU102843  
VERSION AU102843.1 GI:13552364  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki.Y., Tsunoda.T., Taira.H., Mizushima-Sugano.J., Sese.J., Hata  
H., Ota.T., Isogai.T., Tanaka.T., Nakamura.Y., Morishita.S., Okubo  
K., Suyama.A. and Sugano.S.  
TITLE Fine Structural analysis of transcription start sites of human  
JOURNAL mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
COMMENT Unpublished (2001)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp  
Suzuki.Y., Yoshitomo-Nakagawa.K., Maruyama.K., Suyama.A. and Sugano  
S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES  
source  
1..50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CAS11158"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 21 a 11 c 11 g 7 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 107; Length 50;  
Best Local Similarity 77.8%; Pred. No. 7.2e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 agatgctaggtagctgtc 20  
||||| ||||| |||||

Db 48 AGAGGTTGGTATCTGCC 31

RESULT 13  
LOCUS AU102844/c 50 bp mRNA EST 05-APR-2001  
DEFINITION AU102844 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
CAS11168, mRNA sequence.  
ACCESSION AU102844  
VERSION AU102844.1 GI:13552365  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki.Y., Tsunoda.T., Taira.H., Mizushima-Sugano.J., Sese.J., Hata  
H., Ota.T., Isogai.T., Tanaka.T., Nakamura.Y., Morishita.S., Okubo  
K., Suyama.A. and Sugano.S.  
TITLE Fine Structural analysis of transcription start sites of human  
JOURNAL mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
COMMENT Unpublished (2001)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp  
Suzuki.Y., Yoshitomo-Nakagawa.K., Maruyama.K., Suyama.A. and Sugano  
S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES  
source  
1..50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CAS11168"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 21 a 11 c 11 g 7 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 107; Length 50;  
Best Local Similarity 77.8%; Pred. No. 7.2e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 agatgctaggtagctgtc 20  
||||| ||||| |||||

Db 48 AGAGGTTGGTATCTGCC 31

RESULT 14  
LOCUS AA275379 36 bp mRNA EST 31-MAR-1997  
DEFINITION vc10d01.r1 Barstead MPLRBI Mus musculus cDNA clone IMAGE:766081 5'  
similar to gb:J33878 Mus musculus renal sodium phosphate (MOUSE);  
mRNA sequence.  
ACCESSION AA275379  
VERSION AA275379.1 GI:1915865  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 36)  
AUTHORS Marra.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T.,  
Geisler.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M.,  
Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B.,  
Theising.B., Wylie.T., Lennon.G., Soares.B., Willson.R. and  
Waterston.R.  
TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project

WashU-HMNI Mouse EST Project  
Washington University School  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:467001

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 31.

#### FEATURES

source  
1. 36  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
/clone="IMAGE:766081"  
/clone\_lib="Barstead MPLRBL"  
/sex="mixed"  
/tissue\_type="Kidney"  
/dev\_stage="6 weeks"  
/lab\_host="DH108"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACGAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[CATGATTCGGTACC], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT73 vector.  
Library constructed by Bob Barstead."

BASE COUNT 13 a 5 c 10 g .8 t  
ORIGIN

Query Match 57.0%; Score 11.4; DB 4; Length 36;

Best Local Similarity 92.3%; Pred. No. 8.7e+04; Indels 0; Gaps 0;

Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tagatgctagta 14

||| |||||

Db 20 TACATGCTAGTA 32

#### RESULT 15

AZ423252

LOCUS

DEFINITION

AZ423252 39 bp DNA GSS 03-OCT-2000

1M0202H18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0202H18 F, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 39)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0202 row: H column: 18

Seq primer: CGTTGTAACAGCAGCGCCACT

Class: plasmid ends

High quality sequence stop: 39.

#### FEATURES

source  
1. 39  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0202H18"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F."  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWB42 (gil14732114[gb|AF129072.1]), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 9 a 8 c 3 g 19 t  
ORIGIN

Query Match 57.0%; Score 11.4; DB 243; Length 39;

Best Local Similarity 92.3%; Pred. No. 8.8e+04;

Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ctatgtatctgtc 20

||| |||||

Db 1 CTATGTATCTGTC 13

Search completed: October 2, 2001, 15:00:53

Job time: 10822 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:03:47 ; Search time 417.38 Seconds  
(without alignments)  
9.071 Million cell updates/sec

Title: US-09-757-100B-8

Perfect score: 20

Sequence: 1 ctatagctaggtagtctgtc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 460742

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	20	100.0	20	3	US-09-377-310-8
2	15	75.0	15	3	US-09-377-310-28
3	15	75.0	20	3	US-09-377-310-9
C 4	13.4	67.0	43	2	US-08-850-049-17
C 5	13.4	67.0	43	2	US-08-050-478-17
C 6	13.4	67.0	43	4	US-07-858-747B-17
C 7	12.8	64.0	39	4	US-09-342-749-15
8	12.6	63.0	21	1	US-08-474-177-34
9	12.6	63.0	21	1	US-08-487-033-34
10	12.6	63.0	21	1	US-08-480-810-34
11	12.6	63.0	21	2	US-08-508-735-34
12	12.6	63.0	21	2	US-08-848-251-34
13	12.6	63.0	21	2	US-08-486-047-34
14	12.6	63.0	21	3	US-09-120-130-34
15	12.6	63.0	21	3	US-09-115-252-34
16	12.6	63.0	21	3	US-08-986-515-34
17	12.6	63.0	21	3	US-09-120-128-34
18	12.6	63.0	21	4	US-09-120-129-34
19	12.6	63.0	21	4	US-09-201-139-34
20	12.6	63.0	21	4	US-09-120-131-34
21	12.6	63.0	27	1	US-08-309-644A-6
C 22	12.6	63.0	27	1	US-08-309-644A-7
C 23	12.4	62.0	35	1	US-08-810-116-2
24	12.4	62.0	35	2	US-07-930-548A-2
25	12.2	61.0	42	2	US-08-790-963-2
26	12.2	61.0	42	4	US-09-371-774-2
C 27	12.2	61.0	44	1	US-08-168-917-10

C 28	12.2	61.0	44	2	US-08-460-510-10	Sequence 10, Appl
C 29	12.2	61.0	44	2	US-08-460-490-10	Sequence 10, Appl
C 30	12.2	61.0	44	5	PCT-US92-00730-10	Sequence 10, Appl
C 31	12.2	61.0	44	5	PCT-US92-10430-5	Sequence 5, Appl
C 32	12	60.0	15	3	US-09-377-310-29	Sequence 29, Appl
C 33	11.8	59.0	22	2	US-08-267-803B-21	Sequence 21, Appl
C 34	11.6	58.0	20	1	US-07-977-284A-74	Sequence 74, Appl
C 35	11.6	58.0	20	1	US-08-271-880A-172	Sequence 172, App
C 36	11.6	58.0	20	1	US-08-271-880A-189	Sequence 189, App
C 37	11.6	58.0	20	2	US-08-256-426B-74	Sequence 74, Appl
C 38	11.6	58.0	20	2	US-08-910-408-172	Sequence 172, App
C 39	11.6	58.0	20	2	US-08-910-408-189	Sequence 189, App
C 40	11.6	58.0	20	3	US-09-249-215-172	Sequence 172, App
C 41	11.6	58.0	20	3	US-09-249-215-189	Sequence 189, App
C 42	11.6	58.0	40	1	US-07-832-905B-69	Sequence 69, Appl
C 43	11.6	58.0	40	2	US-08-700-757-69	Sequence 69, Appl
C 44	11.2	56.0	22	4	US-08-876-885-21	Sequence 21, Appl
C 45	11.2	56.0	23	1	US-08-464-531-59	Sequence 59, Appl

ALIGNMENTS

RESULT 1  
US-09-377-310-8  
; Sequence 8, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-8

Query Match 100.0%; Score 20; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.049;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatagctaggtagtctgtc 20  
Db 1 ctatagctaggtagtctgtc 20

RESULT 2  
US-09-377-310-28  
; Sequence 28, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

RESULT 5  
US-08-050-478-17/c  
; Sequence 17, Application US/08050478  
; Patent No. 5972596  
; GENERAL INFORMATION:  
; APPLICANT:  
; APPLICANT:  
; APPLICANT:  
; TITLE OF INVENTION: METHOD OF ELIMINATING  
; TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF MRNA  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/050,478  
; FILING DATE: 26-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/02908  
; FILING DATE: 29-MAR-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/858,747  
; FILING DATE: 27-MAR-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: MORRY, MARY J.  
REGISTRATION NUMBER: 34,398  
REFERENCE/DOCKET NUMBER: 2026-4006US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)758-4800  
TELEFAX: (212)751-6849  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
US-08-050-478-17

Query Match 67.0%; Score 13.4; DB 2; Length 43;  
Best Local Similarity 93.3%; Pred. No. 1.5e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gatgctaggtagtctg 18  
||| ||||| ||||| |||||  
Db 19 GAAGCTAGGTATCTG 5

RESULT 6  
US-07-858-747B-17/c  
Sequence 17, Application US/07858747B  
Patent No. 6174666  
GENERAL INFORMATION:  
APPLICANT: PAVLAKIS, GEORGE N.; FELBER, BARBARA  
APPLICANT: K.  
TITLE OF INVENTION: METHOD OF ELIMINATING  
TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF MRNA  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/858,747B  
FILING DATE: 19920327  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: MARY J. MORRY  
REGISTRATION NUMBER: 34,398  
REFERENCE/DOCKET NUMBER: 2026-4006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)758-4800  
TELEFAX: (212)751-6849  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: UNKNOWN  
MOLECULE TYPE: OTHER  
HYPOTHETICAL: YES  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: OLIGONUCLEOTIDE FOR MUTATING NT

OTHER INFORMATION: 3392-3434 OF C-FOS, HUMCFOT, ACC #V01512  
US-07-858-747B-17

Query Match 67.0%; Score 13.4; DB 4; Length 43;  
Best Local Similarity 93.3%; Pred. No. 1.5e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gatgctaggtagtctg 18  
||| ||||| ||||| |||||  
Db 19 GAAGCTAGGTATCTG 5

RESULT 7  
US-09-342-749-15/C  
Sequence 15, Application US/09342749  
Patent No. 6166194  
GENERAL INFORMATION:  
APPLICANT: Wong, Alexander K.C.  
APPLICANT: Tavtigian, Sean V.  
APPLICANT: Teng, David H.-F.  
APPLICANT: Myriad Genetics, Inc.  
TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor  
FILE REFERENCE: 2318-202  
CURRENT APPLICATION NUMBER: US/09/342,749  
CURRENT FILING DATE: 1999-06-29  
EARLIER APPLICATION NUMBER: US 60/091,044  
EARLIER FILING DATE: 1998-06-29  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 15  
LENGTH: 39  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-342-749-15

Query Match 64.0%; Score 12.8; DB 4; Length 39;  
Best Local Similarity 87.5%; Pred. No. 3e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 atgctaggtagtctg 20  
||| ||||| ||||| |||||  
Db 30 ATGATAGTATCCGTC 15

RESULT 8  
US-08-474-177-34  
Sequence 34, Application US/08474177  
Patent No. 5624819  
GENERAL INFORMATION:  
APPLICANT: Skolnick, Mark H.  
APPLICANT: Cannon-Albright, Lisa A.  
APPLICANT: Kamb, Alexander  
TITLE OF INVENTION: GERMLINE MUTATIONS IN THE MTS GENE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,177  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/03537  
FILING DATE: 17-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,938  
FILING DATE: 01-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/215,087  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/215,086  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,369  
FILING DATE: 14-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/214,582  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109348-E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-8300  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: YES  
US-08-474-177-34

Query Match 63.0%; Score 12.6; DB 1; Length 21;  
Best Local Similarity 78.9%; Pred. No. 3.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctatagcttagtctgt 19  
||||| ||| |||||  
Db 3 CTAGAGCGCAATTATCTGT 21

RESULT 9  
US-08-487-033-34  
Sequence 34; Application US/08487033  
Patent No. 5739027  
GENERAL INFORMATION:  
APPLICANT: Ramb, Alexander  
TITLE OF INVENTION: MTS1E1-Beta GENE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,033  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/03316  
FILING DATE: 17-MAR-1995  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/251,938  
FILING DATE: 01-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/215,087  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/215,086  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,369  
FILING DATE: 14-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/214,582  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109348-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-8300  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: YES  
US-08-487-033-34

Query Match 63.0%; Score 12.6; DB 1; Length 21;  
Best Local Similarity 78.9%; Pred. No. 3.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctatagcttagtctgt 19  
||||| ||| |||||  
Db 3 CTAGAGCGCAATTATCTGT 21

RESULT 10  
US-08-480-810-34  
Sequence 34; Application US/08480810  
Patent No. 5801236  
GENERAL INFORMATION:  
APPLICANT: Ramb, Alexander  
TITLE OF INVENTION: MTS1 GENE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,810  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/03316  
FILING DATE: 17-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,938  
FILING DATE: 01-JUN-1994  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/215,087  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/215,086  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,369  
FILING DATE: 14-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/214,582  
FILING DATE: 18-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109348  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: YES  
US-08-480-810-34

Query Match 63.0%; Score 12.6; DB 1; Length 21;  
Best Local Similarity 78.9%; Pred. No. 3.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctatagctcgtagggtatctgt 19  
||||| || | |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

RESULT 11  
US-08-508-735-34  
Sequence 34, Application US/08508735  
Patent No. 5843756  
GENERAL INFORMATION:  
APPLICANT: Stone, Steven  
APPLICANT: Jiang, Ping  
APPLICANT: Kamb, Alexander  
TITLE OF INVENTION: MTS GENE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/508,735  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US to be assigned  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/03316  
FILING DATE: 17-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.

REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109348  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4848  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: YES  
US-08-508-735-34

Query Match 63.0%; Score 12.6; DB 2; Length 21;  
Best Local Similarity 78.9%; Pred. No. 3.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctatagctcgtagggtatctgt 19  
||||| || | |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

RESULT 12  
US-08-848-251-34  
Sequence 34, Application US/08848251  
Patent No. 5989815  
GENERAL INFORMATION:  
APPLICANT: Skolnick, Mark H.  
APPLICANT: Cannon-Albright, Lisa A.  
APPLICANT: Kamb, Alexander  
TITLE OF INVENTION: GERMLINE MUTATIONS IN THE MTS GENE AND  
TITLE OF INVENTION: METHOD FOR DETECTING PREDISPOSITION TO CANCER AT THE MTS  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/848,251  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,083  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: PCT/US95/03537  
FILING DATE: 17-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,938  
FILING DATE: 01-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/215,087  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/215,086  
FILING DATE: 18-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,369  
FILING DATE: 14-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/214,582

; FILING DATE: 18-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109348-G  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
US-08-848-251-34

Query Match 63.0%; Score 12.6; DB 2; Length 21;  
Best Local Similarity 78.9%; Pred. NO. 3.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ctatagctaggtatctgt 19  
||||| ||| | |||||  
Db 3 CTAGAGCGCAATTATCTGT 21

RESULT 13  
US-08-486-047-34  
; Sequence 34, Application US/08486047  
; Patent No. 5994095  
; GENERAL INFORMATION:  
; APPLICANT: Kamb, Alexander  
; TITLE OF INVENTION: MTS2 GENE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486.047  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/03316  
; FILING DATE: 17-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,938  
; FILING DATE: 01-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,087  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,086  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/227,369  
; FILING DATE: 14-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/214,582  
; FILING DATE: 18-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.

; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109348-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
US-08-486-047-34

Query Match 63.0%; Score 12.6; DB 2; Length 21;  
Best Local Similarity 78.9%; Pred. NO. 3.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ctatagctaggtatctgt 19  
||||| ||| | |||||  
Db 3 CTAGAGCGCAATTATCTGT 21

RESULT 14  
US-09-120-130-34  
; Sequence 34, Application US/09120130  
; Patent No. 6037462  
; GENERAL INFORMATION:  
; APPLICANT: Kamb, Alexander  
; TITLE OF INVENTION: MTS1 GENE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/120,130  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/480,810  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,938  
; FILING DATE: 01-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,087  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,086  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/227,369  
; FILING DATE: 14-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/214,582  
; FILING DATE: 18-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109348  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
US-09-120-130-34

Query Match 63.0% Score 12.6; DB 3; Length 21;  
Best Local Similarity 78.9%; Pred. No. 3.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgt 19  
||||| ||| |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

RESULT 15  
US-09-115-252-34  
; Sequence 34, Application US/09115252  
; Patent No. 6060301  
; GENERAL INFORMATION:  
; APPLICANT: Kamb, Alexander  
; TITLE OF INVENTION: MTS1 GENE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/115,252  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,810  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: PCT/US95/03316  
; FILING DATE: 17-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,938  
; FILING DATE: 01-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,087  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,086  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/227,369  
; FILING DATE: 14-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/214,582  
; FILING DATE: 18-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-109348  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810

; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
US-09-115-252-34

Query Match 63.0% Score 12.6; DB 3; Length 21;  
Best Local Similarity 78.9%; Pred. No. 3.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgt 19  
||||| ||| |||||  
Db 3 CTAGAGCGGAATTATCTGT 21

Search completed: October 2, 2001, 16:03:47  
Job time: 14591 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:42 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-9  
Perfect score: 20  
Sequence: 1 ttctgtagatgctaggtat 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_pl1: \*  
13: gb\_pl2: \*  
14: gb\_pl3: \*  
15: gb\_pl4: \*  
16: em\_ba1: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rod: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rod: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pl: \*  
48: em\_rod: \*  
49: em\_sts: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_vi: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_vil: \*  
59: gb\_vil2: \*  
60: gb\_htg1: \*  
61: gb\_htg2: \*  
62: gb\_htg3: \*  
63: gb\_htg4: \*  
64: gb\_htg5: \*  
65: gb\_htg6: \*  
66: gb\_htg7: \*  
67: gb\_htg8: \*  
68: gb\_htg9: \*  
69: gb\_htg10: \*  
70: gb\_htg11: \*  
71: gb\_htg12: \*  
72: gb\_htg13: \*  
73: gb\_htg14: \*  
74: gb\_htg15: \*  
75: gb\_htg16: \*  
76: gb\_htg17: \*  
77: gb\_htg18: \*  
78: gb\_htg19: \*  
79: gb\_htg20: \*  
80: gb\_htg21: \*  
81: gb\_htg22: \*  
82: gb\_htg23: \*  
83: gb\_htg24: \*  
84: gb\_htg25: \*  
85: gb\_pr1: \*  
86: gb\_pr2: \*  
87: gb\_pr3: \*  
88: gb\_pr4: \*  
89: gb\_pr5: \*  
90: gb\_pr6: \*  
91: gb\_pr7: \*  
92: gb\_pr8: \*  
93: gb\_pr9: \*  
94: gb\_rol: \*  
95: gb\_rod: \*  
96: gb\_in4: \*  
97: gb\_pr10: \*  
98: em\_ba3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	13.6	68.0	31	10	I24252
c 2	12.4	62.0	37	10	AX100491
c 3	12.2	61.0	38	9	A25775
4	12.2	61.0	41	9	AR034285
5	12.2	61.0	41	9	AR035391
6	12.2	61.0	41	9	AR050806
7	12.2	61.0	41	9	AR053811
8	12.2	61.0	41	9	AR091592
					I24252 Sequence 39
					AX100491 Sequence
					A25775 antise o
					AR034285 Sequence
					AR035391 Sequence
					AR050806 Sequence
					AR053811 Sequence
					AR091592 Sequence

```

9      12 60.0 26 9 A83432
10     12 60.0 33 9 AX019713
11     12 60.0 35 9 AR030753
12     12 60.0 35 9 AR075961
13     12 60.0 39 9 A42732
14     12 60.0 39 10 187238
15     12 60.0 39 10 187238
16     12 60.0 44 10 190213
17     12 60.0 45 97 HUMA1DSC2
18     11.8 59.0 33 9 A69048
19     11.8 59.0 36 9 A71632
20     11.8 59.0 40 9 A43390
21     11.8 59.0 42 9 AR054989
22     11.6 58.0 37 10 AX097685
23     11.6 58.0 50 9 AR040802
24     11.4 57.0 28 9 A98431
25     11.4 57.0 33 9 AX021032
26     11.4 57.0 45 10 AX099973
27     11.2 56.0 21 10 AX097355
28     11.2 56.0 23 10 130679
29     11.2 56.0 23 10 146138
30     11.2 56.0 25 9 AX043430
31     11.2 56.0 25 10 E50316
32     11.2 56.0 30 9 A87635
33     11.2 56.0 31 9 AR029080
34     11.2 56.0 40 9 A17046
35     11.2 56.0 40 9 A17446
36     11.2 56.0 40 9 AR014421
37     11.2 56.0 40 10 111864
38     11.2 56.0 41 9 A47211
39     11.2 56.0 43 10 E21659
40     11.2 56.0 48 97 S76309
41     11 55.0 20 10 E07210
42     11 55.0 21 9 A18195
43     11 55.0 21 9 A22305
44     11 55.0 22 9 AR053115
45     11 55.0 24 9 A22306

```

## ALIGNMENTS

```

RESULT 1
124252 LOCUS 124252 31 bp DNA PAT 07-OCT-1996
DEFINITION Sequence 39 from patent US 5543293.
ACCESSION 124252
VERSION 124252.1 GI:1604122
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 31)
AUTHORS Gold,L. and Tasset,D.
TITLE DNA ligands of thrombin
JOURNAL Patent: US 5543293-A 39 06-AUG-1996;
FEATURES Location/Qualifiers
source 1..31
BASE COUNT 5 a 3 c 11 g 12 t
ORIGIN

Query Match 68.0%; Score 13.6; DB 10; Length 31;
Best Local Similarity 80.0%; Pred. No. 1.1e+04;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 ttttgcctagctaggtat 20
      ||||| || ||||| ||
Db 2 TTTTGGTATAGCTAGGTGT 21

```

RESULT 2

```

AX100491/c
LOCUS AX100491 37 bp DNA PAT 10-APR-2001
DEFINITION Sequence 5 from Patent WO0121783.
ACCESSION AX100491
VERSION AX100491.1 GI:13619505
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE 1 (bases 1 to 37)
AUTHORS Jiao,S., Habben,J.E. and Niu,X.
TITLE Seed-preferred promoter from maize
JOURNAL Patent: WO 0121783-A 5 29-MAR-2001;
PIONEER HT-BRED INTERNATIONAL, INC. (US)
FEATURES Location/Qualifiers
source 1..37
/organism="Zea mays"
/db_xref="taxon:4577"
BASE COUNT 9 a 10 c 7 g 11 t
ORIGIN

Query Match 62.0%; Score 12.4; DB 10; Length 37;
Best Local Similarity 92.9%; Pred. No. 4.8e+04;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 6 ctatagctagtgta 19  
 ||||| |||||  
 Db 37 CTAGATAGTAGGTA 24

```

RESULT 3
A25775/c LOCUS A25775 38 bp DNA PAT 22-MAR-1995
DEFINITION antisense oligonucleotide OWB22.
ACCESSION A25775
VERSION A25775.1 GI:904750
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 38)
AUTHORS
TITLE BIOCIDAL PROTEINS
JOURNAL Patent: WO 9304586-A 8 18-MAR-1993;
FEATURES Location/Qualifiers
source 1..38
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 11 a 8 c 7 g 12 t
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 38;
Best Local Similarity 82.4%; Pred. No. 6.1e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 4 tgctagatgctaggtat 20  
 ||||| ||||| |||||  
 Db 26 TGATAGAGCTAGCTAT 10

```

RESULT 4
AR034285 LOCUS AR034285 41 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 8 from patent US 5869337.
ACCESSION AR034285
VERSION AR034285.1 GI:5949890
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

```

```
Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Crabtree,G.R., Schreiber,S.L., Spencer,D.M., Wandless,T.J. and
        Belshaw,P.
TITLE Regulated transcription of targeted genes and other biological
        events
JOURNAL Patent: US 5869337-A 8 09-FEB-1999;
FEATURES Location/Qualifiers
        source 1..41
BASE COUNT 7 a 13 c 9 g 12 t
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 41;
Best Local Similarity 82.4%; Pred. No. 6.1e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgtctagatgctaggt 18
    | ||||| |||||
Db 13 TCTGCTACTTGTCTAGGT 29

RESULT 5
AR035391
LOCUS AR035391 41 bp DNA 29-SEP-1999
DEFINITION Sequence 8 from patent US 5871753.
ACCESSION AR035391
VERSION AR035391.1 GI:5952059
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Crabtree,G.R., Schreiber,S.L., Spencer,D.M., Wandless,T.J.,
        Belshaw,P. and Ho,S.
TITLE Regulated transcription of targeted genes and other biological
        events
JOURNAL Patent: US 5871753-A 8 16-FEB-1999;
FEATURES Location/Qualifiers
        source 1..41
BASE COUNT 7 a 13 c 9 g 12 t
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 41;
Best Local Similarity 82.4%; Pred. No. 6.1e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgtctagatgctaggt 18
    | ||||| |||||
Db 13 TCTGCTACTTGTCTAGGT 29

RESULT 6
AR050806
LOCUS AR050806 41 bp DNA 29-SEP-1999
DEFINITION Sequence 8 from patent US 5830462.
ACCESSION AR050806
VERSION AR050806.1 GI:5974170
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Crabtree,G.R., Schreiber,S.L., Spencer,D.M., Wandless,T.J. and
        Belshaw,P.
TITLE Regulated transcription of targeted genes and other biological
        events
JOURNAL Patent: US 5830462-A 8 03-NOV-1998;
FEATURES Location/Qualifiers
        source 1..41

Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Crabtree,G.R., Schreiber,S.L., Spencer,D.M., Wandless,T.J. and
        Belshaw,P.
TITLE Regulated transcription of targeted genes and other biological
        events
JOURNAL Patent: US 5869337-A 8 09-FEB-1999;
FEATURES Location/Qualifiers
        source 1..41
BASE COUNT 7 a 13 c 9 g 12 t
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 41;
Best Local Similarity 82.4%; Pred. No. 6.1e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgtctagatgctaggt 18
    | ||||| |||||
Db 13 TCTGCTACTTGTCTAGGT 29

RESULT 7
AR053811
LOCUS AR053811 41 bp DNA 29-SEP-1999
DEFINITION Sequence 8 from patent US 5834266.
ACCESSION AR053811
VERSION AR053811.1 GI:5978673
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Crabtree,G.R., Schreiber,S.L., Spencer,D.M., Wandless,T.J. and
        Belshaw,P.
TITLE Regulated apoptosis
JOURNAL Patent: US 5834266-A 8 10-NOV-1998;
FEATURES Location/Qualifiers
        source 1..41
BASE COUNT 7 a 13 c 9 g 12 t
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 41;
Best Local Similarity 82.4%; Pred. No. 6.1e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgtctagatgctaggt 18
    | ||||| |||||
Db 13 TCTGCTACTTGTCTAGGT 29

RESULT 8
AR091592
LOCUS AR091592 41 bp DNA 07-SEP-2000
DEFINITION Sequence 8 from patent US 5994313.
ACCESSION AR091592
VERSION AR091592.1 GI:10018347
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Crabtree,G.R., Schreiber,S.L., Spencer,D.M., Wandless,T.J. and
        Belshaw,P.
TITLE Regulated apoptosis
JOURNAL Patent: US 5994313-A 8 30-NOV-1999;
FEATURES Location/Qualifiers
        source 1..41
BASE COUNT 7 a 13 c 9 g 12 t
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 41;
Best Local Similarity 82.4%; Pred. No. 6.1e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgtctagatgctaggt 18
    | ||||| |||||
Db 13 TCTGCTACTTGTCTAGGT 29

RESULT 9
AR091592
LOCUS AR091592 41 bp DNA 07-SEP-2000
DEFINITION Sequence 8 from patent US 5994313.
ACCESSION AR091592
VERSION AR091592.1 GI:10018347
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Crabtree,G.R., Schreiber,S.L., Spencer,D.M., Wandless,T.J. and
        Belshaw,P.
TITLE Regulated apoptosis
JOURNAL Patent: US 5994313-A 8 30-NOV-1999;
FEATURES Location/Qualifiers
        source 1..41
BASE COUNT 7 a 13 c 9 g 12 t
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 41;
Best Local Similarity 82.4%; Pred. No. 6.1e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgtctagatgctaggt 18
    | ||||| |||||
Db 13 TCTGCTACTTGTCTAGGT 29
```

```

Db 13 TCTGCTACTGCTAGCT 29

RESULT 9
LOCUS A83432 26 bp DNA PAT 21-JAN-2000
DEFINITION Sequence 18 from Patent WO9850067.
ACCESSION A83432
VERSION A83432.1 GI:6732770
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 26)
AUTHORS Goebel,W. and Demuth,A.
TITLE USE OR A SECRETION VECTOR FOR FERTILITY CONTROL BY ORAL VACCINATION
JOURNAL Patent: WO 9850067-A 18 12-NOV-1998;
GOEBEL WERNER (DE); SCHERING AG (DE)
FEATURES
source
1..26
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 6 a 5 c 7 g 8 t
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 26;
Best Local Similarity 100.0%; Pred. No. 8e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtctagatgc 13
|||||
Db 4 TTTGCTAGATGC 15

RESULT 10
AX019713
LOCUS AX019713 33 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 2 from Patent WO9937807.
ACCESSION AX019713
VERSION AX019713.1 GI:10043553
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 33)
AUTHORS Hentze,M.W. and Paraskeva,E.
TITLE Method of isolation of rna-binding compounds
JOURNAL Patent: WO 9937807-A 2 29-JUL-1999;
EUROP MOLECULAR BIOLOGY LAB (DE); HENTZE MATTHIAS W (DE); PARASKEVA
EFROSYNI (DE)
FEATURES
source
1..33
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer for amplification by PCR"
BASE COUNT 7 a 4 c 4 g 18 t
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 33;
Best Local Similarity 75.0%; Pred. No. 7.9e+04;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttgtctagatgc 20
|||||
Db 2 TTTTCTAGATTATGTTAT 21

RESULT 11
AX030753
LOCUS AX030753 35 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5861301.
ACCESSION AX030753
VERSION AX030753.1 GI:5943967
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Terman,B.Israel and Carrion,M.Eduardo.
TITLE Recombinant kinase insert domain containing receptor and gene
encoding same
JOURNAL Patent: US 5861301-A 2 19-JAN-1999;
LOCATION/Qualifiers
1..35
/organism="unknown"
BASE COUNT 5 a 9 c 8 g 8 t 5 others
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 35;
Best Local Similarity 63.2%; Pred. No. 7.9e+04;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtctagatgc 20
|||||
Db 12 KTTNCTRGCGCAGTCT 30

RESULT 12
AR075961
LOCUS AR075961 35 bp DNA PAT 30-AUG-2000
DEFINITION Sequence 22 from patent US 5958713.
ACCESSION AR075961
VERSION AR075961.1 GI:10002707
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Thastrup,O., Tullin,S.slashed.ren, Poulsen,L.Kongsbak and Bj.o
slashed.rp.S.Petersen.
TITLE Method of detecting biologically active substances by using green
fluorescent protein
JOURNAL Patent: US 5958713-A 22 28-SEP-1999;
LOCATION/Qualifiers
1..35
/organism="unknown"
BASE COUNT 16 a 7 c 3 g 9 t
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 35;
Best Local Similarity 75.0%; Pred. No. 7.9e+04;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttgtctagatgc 20
|||||
Db 21 TTTTGAAGATCTTTGGTAT 2

RESULT 13
A42732/c
LOCUS A42732 39 bp DNA PAT 06-MAR-1997
DEFINITION Sequence 64 from Patent WO9503412.
ACCESSION A42732
VERSION A42732.1 GI:2298181
KEYWORDS
SOURCE Mycobacterium fortuitum.
ORGANISM Mycobacterium fortuitum.
REFERENCE 1 (bases 1 to 39)
AUTHORS Mabilat,C. and Christen,R.

```

TITLE NUCLEOTIDE FRAGMENT OF MYCOBACTERIAL RIBOSOMAL RNA 23S, PROBES AND PRIMERS DERIVED THEREFROM, REAGENT AND METHOD FOR DETECTING SAID FRAGMENT

JOURNAL Patent: WO 9503412-A 64 02-FEB-1995;

BIO MERIEUX (FR)

COMMENT Other publication CA 2145172 950202

Other publication FR 2709310 950303.

FEATURES Location/Qualifiers

source 1. .39

/organism="Mycobacterium fortuitum"

/strain="CIP 140 410 001"

/db\_xref="taxon:1786"

BASE COUNT 12 a 11 c 9 g 7 t

ORIGIN

Query Match

Best Local Similarity 60.0%; Score 12; DB 9; Length 39;

Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttttgctagatgctaggtat 20

||||| ||||| |||||

Db 25 TCTTGGTAGATGTCAGGTTT 6

RESULT 14

187238/c

LOCUS

187238

DEFINITION

Sequence 64 from patent US 5703217.

ACCESSION

187238

VERSION

187238.1

KEYWORDS

GI:3206956

SOURCE

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 39)

AUTHORS

Mabilat,C. and Christen,R.

TITLE

Nucleotide fragment of the 23S ribosomal RNA of mycobacteria,

derived probes and primers, reagent and detection method

JOURNAL

Patent: US 5703217-A 64 30-DEC-1997;

FEATURES

Location/Qualifiers

source 1. .39

/organism="unknown"

BASE COUNT 12 a 11 c 9 g 7 t

ORIGIN

Query Match

Best Local Similarity 60.0%; Score 12; DB 10; Length 39;

Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttttgctagatgctaggtat 20

||||| ||||| |||||

Db 25 TCTTGGTAGATGTCAGGTTT 6

RESULT 15

190213/c

LOCUS

190213

DEFINITION

Sequence 39 from patent US 5723595.

ACCESSION

190213

VERSION

190213.1

KEYWORDS

GI:3410153

SOURCE

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 44)

AUTHORS

Thompson,G.A. and Knauf,V.C.

TITLE

Plant desaturases--compositions and uses

JOURNAL

Patent: US 5723595-A 39 03-MAR-1998;

FEATURES

Location/Qualifiers

source 1. .44

/organism="unknown"

BASE COUNT 15 a 9 c 11 g 9 t

ORIGIN

Query Match

Best Local Similarity 60.0%; Score 12; DB 10; Length 44;

Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttttgctagatgctaggtat 20

||||| ||||| |||||

Db 42 TTTTCCTAGTTGTTTACGAT 23

Search completed: October 2, 2001, 15:56:44

Job time: 14168 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:00:53 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-9

Perfect score: 20

Sequence: 1 ttgtgtagatgctaggat 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0066 row: G column: 07  
Seq primer: CGTTGTAAACGACGCGCAGT  
Class: plasmid ends  
High quality sequence stop: 23.  
Location/Qualifiers  
1. 23  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUC1M0066G07"  
/clone\_lib="Mouse 10kb plasmid UUC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gil4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

FEATURES  
source  
1. 23  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUC1M0066G07"  
/clone\_lib="Mouse 10kb plasmid UUC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gil4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 6 a 6 c 4 g 7 t  
ORIGIN  
Query Match 64.0%; Score 12.8; DB 241; Length 23;  
Best Local Similarity 87.5%; Pred. No. 1.9e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ttgtctagatgctagg 17  
|||||||  
Db 19 TTTGCTAGATCCAAAG 4

RESULT 2  
LOCUS AU104064/c  
DEFINITION AU104064 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
HEP13823, mRNA sequence.  
ACCESSION AU104064  
VERSION AU104064.1 GI:13553585  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
K., Suyama,A. and Sugano,S.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
c 1	12.8	64.0	23	241	AZ336049	AM0066G07
c 2	12.6	63.0	50	107	AU104064	AU104064
c 3	12.2	61.0	35	250	A2841152	2M0139006
c 4	12	60.0	49	243	A2421911	1M0200022
c 5	11.8	59.0	37	258	TA254H02Q	AL483435 T. brucei
c 6	11.8	59.0	46	13	AA916137	og32c10.s
c 7	11.6	58.0	37	258	TA383E12Q	BE900863 T. brucei
c 8	11.6	58.0	44	141	BE900863	601673838
c 9	11.6	58.0	50	107	AU104334	AU104334
c 10	11.4	57.0	36	4	AA275379	vc10d01.r
c 11	11.2	56.0	25	243	AZ396383	AM0160720
c 12	11.2	56.0	34	244	A4292829	1M0327J11
c 13	11.2	56.0	39	106	AU006827	AU006827
c 14	11.2	56.0	41	246	A2579388	1M0363P15
c 15	11.2	56.0	47	246	A2588937	1M0397M06
c 16	11.2	56.0	49	11	AA767683	oa45b11.s
c 17	11.2	56.0	50	107	AU102553	AU102553
c 18	11	55.0	36	249	A2805882	2M0067G01
c 19	11	55.0	36	250	A2827044	2M0103A13
c 20	11	55.0	37	244	A2478803	1M0299A08
c 21	11	55.0	40	16	AI146712	qb92a01.x
c 22	11	55.0	41	242	A2358753	1M0101M13
c 23	11	55.0	49	11	AA782046	ai48d07.s
c 24	11	55.0	49	246	A2593107	1M0404Q22
c 25	11	55.0	50	107	AU104063	AU104063
c 26	10.8	54.0	21	246	A2609424	1M0434016
c 27	10.8	54.0	27	247	A2658164	1M0534118
c 28	10.8	54.0	30	165	BE276866	601178428
c 29	10.8	54.0	31	139	BE729154	601561047
c 30	10.8	54.0	33	143	BF026570	601672380
c 31	10.8	54.0	45	244	A467641	1M0279109
c 32	10.8	54.0	47	258	TA388F09Q	AL498213 T. brucei
c 33	10.6	53.0	22	249	A2774731	2M0004M18
c 34	10.6	53.0	25	249	A2773636	2M0001007
c 35	10.6	53.0	30	155	C01989	HUMG000401
c 36	10.6	53.0	31	14	AA981706	ua25605.r
c 37	10.6	53.0	33	249	AZ784432	2M0027E05
c 38	10.6	53.0	34	258	TA174F10Q	AL474538 T. brucei
c 39	10.6	53.0	35	155	C00481	HUMG000800
c 40	10.6	53.0	37	143	BF026747	601671956
c 41	10.6	53.0	38	258	TA50D02Q	AL456084 T. brucei
c 42	10.6	53.0	39	258	TA209A11P	AL479803 T. brucei
c 43	10.6	53.0	40	8	AA531870	AA531870 TgESTz48
c 44	10.6	53.0	42	242	AZ370282	1M0121C16
c 45	10.6	53.0	46	104	A1948986	wq17h03.x

ALIGNMENTS

RESULT 1  
LOCUS AZ336049/c  
DEFINITION AZ336049 Mouse 10kb plasmid UUC1M library Mus musculus genomic  
clone UUC1M0066G07 F, DNA sequence.  
ACCESSION AZ336049  
VERSION AZ336049.1 GI:10404965  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 23)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

TITLE Fine Structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
JOURNAL Unpublished (2001)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano  
, S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

Location/Qualifiers

1..50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HEP13823"

/clone\_lib="Sugano Homo sapiens cDNA library"

8 a 15 c 15 g 12 t

## BASE COUNT

ORIGIN

Query Match 63.0%; Score 12.6; DB 107; Length 50;

Best Local Similarity 78.9%; Pred. No. 2.6e+04;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ttctgtagtagtagta 19

||| ||||| |||||

Db 25 TTCTGCTAGAGGAGGTA 7

## RESULT

AZ841152/c

LOCUS

2M0139D06F Mouse 35 bp DNA GSS 20-FEB-2001

clone UUGC2M0139D06 F, DNA sequence.

ACCESSION

VERSION

AZ841152

KEYWORDS

SS.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 35)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0139 row: D column: 06

Seq primer: CGTTGTAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 35.

Location/Qualifiers

1..35

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0139D06"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and

purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 14; Conservative

0; Mismatches 3; Indels

0; Gaps 0;

Oy 1 ttctgtagtagtagg 17

||| ||||| |||||

Db 20 TTTAGGTAGATCTTAGG 4

## RESULT

AZ421911/c

LOCUS

IM0200D22F Mouse 49 bp DNA GSS 03-OCT-2000

clone UUGC1M0200D22 F, DNA sequence.

ACCESSION

VERSION

AZ421911

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 49)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0200 row: D column: 22

Seq primer: CGTTGTAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 49.

Location/Qualifiers

1..49

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0200D22"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gill4732114(gb)/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 11 a 9 c 10 g 19 t  
ORIGIN

Query Match 60.0%; Score 12; DB 243; Length 49;  
Best Local Similarity 75.0%; Pred. No. 5.3e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttttgctagctagctagctat 20  
||| | |||| |||| ||  
Db 36 TTTGACAAGATCCTAGGAAT 17

RESULT 5  
TA254H02Q 37 bp DNA GSS 13-DEC-2000  
LOCUS T. brucei sheared genomic DNA clone 254h02, reverse sequence,  
genomic survey sequence.  
ACCESSION AL483435.1 GI:11849761  
VERSION AL483435.1  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 37)  
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,  
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,  
Melville,S.E., Rajandream,M.A. and Barrell,B.G.  
TITLE Direct Submission  
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
nh@sanger.ac.uk  
COMMENT Constructed at the Institute for Genomic Research (TIGR),  
Rockville, MD. Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
to give a tight size distribution (4 kb). The v + i method used for the library construction is  
described in detail in Smith, H. and Venter, J.C. (Making small  
insert libraries for whole genome shotgun sequencing projects. In  
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999).  
Email: nelsayed@tigr.org  
Details of T. brucei sequencing at the Sanger Centre are available  
at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

FEATURES  
source 1. .37  
/organism="Trypanosoma brucei"  
/strain="TREU927"  
/db\_xref="taxon:5691"  
/clone="254h02"  
BASE COUNT 8 a 8 c 10 g 11 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 258; Length 37;  
Best Local Similarity 86.7%; Pred. No. 6.5e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tttgctagctagctag 16  
||| |||| |||| ||  
Db 3 TTTTCTAGGTCTAG 17

RESULT 6  
AA916137 46 bp mRNA EST 14-APR-1998  
LOCUS OQ32cl0.s1 NCI-CGAP Br7 Homo sapiens cDNA clone IMAGE:1441554.3'  
DEFINITION Similar to SW:NU2M\_HUMAN P03891 NADH-UBIQUINONE OXIDOREDUCTASE  
CHAIN 2;; mRNA sequence.  
ACCESSION AA916137  
VERSION AA916137.1 GI:3055529  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 46)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
unknown library type

Trace considered overall poor quality  
Seq primer: -40ml3 fwd. EF from Amerisham  
High quality sequence stop: 1.  
FEATURES  
Location/Qualifiers  
source 1. .46  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1441554"  
/clone\_lib="NCI-CGAP\_Br7"  
/lab\_host="DH10B"  
/note="Organ: breast; Vector: pCMV-SPORT4; Site:1; SalI;  
Site:2; NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.2 kb. Life Technologies catalog  
# :10985-018"

BASE COUNT 12 a 4 c 19 g 11 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 13; Length 46;  
Best Local Similarity 86.7%; Pred. No. 6.7e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 gctagctagctagcta 19  
||||| | |||| ||  
Db 2 GCTAGCTCTAGGTA 16

RESULT 7  
TA383E12Q 37 bp DNA GSS 13-DEC-2000  
LOCUS T. brucei sheared genomic DNA clone 383el2, reverse sequence,  
genomic survey sequence.  
ACCESSION AL497988  
VERSION AL497988.1 GI:11873710  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
REFERENCE 1 (bases 1 to 37)  
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,



```

KEYWORDS EST.
SOURCE house musculus
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 36)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:467001
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 31.
Location/Qualifiers
1. .36
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="IMAGE:766081"
/clone_lib="Barstead MPLRBI"
/sex="mixed"
/tissue_type="Kidney"
/dev_stage="6 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(GT) primer [5',
TGTAGCAATCTGAAGTGGGCGCGCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(CATGATTCGGACCC), digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT7T3 vector.
Library constructed by Bob Barstead."
BASE COUNT 13 a 5 c 10 g 8 t
ORIGIN

Query Match 57.0%; Score 11.4; DB 4; Length 36;
Best Local Similarity 92.3%; Pred. No. 1e+05;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 tagatgctaggta 19
|||||
Db 20 TACATGCTAGGTA 32

RESULT 11
AZ396383/c 25 bp DNA GSS 03-OCT-2000
LOCUS
DEFINITION IM0160J20R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0160J20 R, DNA sequence.
ACCESSION AZ396383
VERSION AZ396383.1 GI:10511455
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 25)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

TITLE and Wright,D., Weiss,R.
JOURNAL Mouse whole genome scaffolding with paired end reads from 10kb
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: gdunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0160 row: J column: 20
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 25.
Location/Qualifiers
1. .25
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0160J20"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gil47321141gbIAFI29072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 7-a 8 c 6 g 4 t
ORIGIN

Query Match 56.0%; Score 11.2; DB 243; Length 25;
Best Local Similarity 81.2%; Pred. No. 1.2e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 tgcctagatgctaggta 19
|||||
Db 19 TCCTGTATGCCAGGTA 4

RESULT 12
AZ492829 34 bp DNA GSS 05-OCT-2000
LOCUS
DEFINITION IM0327J11F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0327J11 F, DNA sequence.
ACCESSION AZ492829
VERSION AZ492829.1 GI:10665914
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 34)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

```



```

/db_xref="taxon:10090"
/clone="UUGC1M0363P15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      8 a      8 c      9 g      16 t
ORIGIN

Query Match      56.0%; Score 11.2; DB 246; Length 41;
Best Local Similarity 81.2%; Pred. No. 1.3e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttttgcctagatgctag 16
   ||||| ||||| |||||
Db 10 TTCTGCTTGATCTAG 25

RESULT 15
AZ588937/c
LOCUS      47 bp      DNA      GSS      13-DEC-2000
DEFINITION 1M0397M06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0397M06 R, DNA sequence.
ACCESSION  AZ588937
VERSION     AZ588937.1 GI:11711127
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dduwn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0397 row: M column: 06
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 47.
Location/Qualifiers
1. .47
/organism="Mus musculus"
FEATURES
source

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0397M06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      16 a      8 c      11 g      12 t
ORIGIN

```

```

Query Match      56.0%; Score 11.2; DB 246; Length 47;
Best Local Similarity 81.2%; Pred. No. 1.4e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gctagatgctaggtat 20
   ||||| ||||| |||||
Db 20 GCTAGACTAGTGTGT 5

```

Search completed: October 2, 2001, 15:00:54  
Job time: 10823 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:44 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-10

Perfect score: 20

Sequence: 1 ggtaagcagctgccattatt 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_ba3:\*
- 4: gb\_in1:\*
- 5: gb\_in2:\*
- 6: gb\_in3:\*
- 7: gb\_cm:\*
- 8: gb\_ov:\*
- 9: gb\_pat1:\*
- 10: gb\_pat2:\*
- 11: gb\_ph:\*
- 12: gb\_pl1:\*
- 13: gb\_pl2:\*
- 14: gb\_pl3:\*
- 15: gb\_pl4:\*
- 16: em\_ba1:\*
- 17: em\_ba2:\*
- 18: em\_fun:\*
- 19: em\_htgo\_hum:\*
- 20: em\_htgo\_inv:\*
- 21: em\_htgo\_rod:\*
- 22: em\_htg\_hum1:\*
- 23: em\_htg\_hum2:\*
- 24: em\_htg\_hum3:\*
- 25: em\_htg\_hum4:\*
- 26: em\_htg\_hum5:\*
- 27: em\_htg\_hum6:\*
- 28: em\_htg\_hum7:\*
- 29: em\_htg\_hum8:\*
- 30: em\_htg\_inv1:\*
- 31: em\_htg\_inv2:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_rod:\*
- 34: em\_hum1:\*
- 35: em\_hum2:\*
- 36: em\_hum3:\*
- 37: em\_hum4:\*
- 38: em\_hum5:\*
- 39: em\_hum6:\*
- 40: em\_hum7:\*
- 41: em\_in:\*
- 42: em\_cm:\*
- 43: em\_or:\*

- 44: em\_ov:\*
- 45: em\_pat:\*
- 46: em\_ph:\*
- 47: em\_pl:\*
- 48: em\_ro:\*
- 49: em\_sts:\*
- 50: em\_sy:\*
- 51: em\_un:\*
- 52: em\_vi:\*
- 53: gb\_sts1:\*
- 54: gb\_sts2:\*
- 55: gb\_sts3:\*
- 56: gb\_sy:\*
- 57: gb\_un:\*
- 58: gb\_vil:\*
- 59: gb\_vil2:\*
- 60: gb\_htg1:\*
- 61: gb\_htg2:\*
- 62: gb\_htg3:\*
- 63: gb\_htg4:\*
- 64: gb\_htg5:\*
- 65: gb\_htg6:\*
- 66: gb\_htg7:\*
- 67: gb\_htg8:\*
- 68: gb\_htg9:\*
- 69: gb\_htg10:\*
- 70: gb\_htg11:\*
- 71: gb\_htg12:\*
- 72: gb\_htg13:\*
- 73: gb\_htg14:\*
- 74: gb\_htg15:\*
- 75: gb\_htg16:\*
- 76: gb\_htg17:\*
- 77: gb\_htg18:\*
- 78: gb\_htg19:\*
- 79: gb\_htg20:\*
- 80: gb\_htg21:\*
- 81: gb\_htg22:\*
- 82: gb\_htg23:\*
- 83: gb\_htg24:\*
- 84: gb\_htg25:\*
- 85: gb\_pr1:\*
- 86: gb\_pr2:\*
- 87: gb\_pr3:\*
- 88: gb\_pr4:\*
- 89: gb\_pr5:\*
- 90: gb\_pr6:\*
- 91: gb\_pr7:\*
- 92: gb\_pr8:\*
- 93: gb\_pr9:\*
- 94: gb\_rol:\*
- 95: gb\_ro2:\*
- 96: gb\_in4:\*
- 97: gb\_pr10:\*
- 98: em\_ba3:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	13.8	69.0	30	9	AR084719 Sequence
2	13.6	68.0	30	9	A84595 Sequence 5
3	13.6	68.0	30	9	AR079892 Sequence
C 4	13.4	67.0	18	9	A97844 Sequence 12
C 5	13.4	67.0	18	10	I44719 Sequence 45
C 6	13.4	67.0	18	10	I52125 Sequence 45
C 7	12.8	64.0	21	10	AX096725 Sequence
C 8	12.8	64.0	26	10	E04774 Synthetic D

9 12.6 63.0 32 10 106269 106269 Sequence 27  
c 10 12.4 62.0 21 10 AX094845 AX094845 Sequence  
c 11 12.4 62.0 21 10 AX095462 AX095462 Sequence  
c 12 12.4 62.0 39 9 AR032341 AR032341 Sequence  
c 13 12.2 61.0 25 9 AR016058 AR016058 Sequence  
c 14 12.2 61.0 25 9 AR025451 AR025451 Sequence  
c 15 12.2 61.0 25 9 AR075530 AR075530 Sequence  
c 16 12.2 61.0 25 9 AX019566 AX019566 Sequence  
c 17 12.2 61.0 49 9 AR087993 AR087993 Sequence  
c 18 12.2 61.0 25 10 E12395 E12395 PCR primer.  
c 19 12.2 60.0 39 9 AR088301 AR088301 Sequence  
c 20 12.2 60.0 44 10 I13034 I13034 Sequence  
c 21 12.2 60.0 48 10 I13041 I13041 Sequence  
c 22 12.2 60.0 48 10 I13042 I13042 Sequence  
c 23 11.8 59.0 18 9 A97839 A97839 Sequence  
c 24 11.8 59.0 18 9 A97845 A97845 Sequence  
c 25 11.8 59.0 22 9 A97943 A97943 Sequence  
c 26 11.8 59.0 39 9 AR080603 AR080603 Sequence  
c 27 11.8 59.0 41 9 A81398 A81398 Sequence  
c 28 11.8 59.0 41 9 A82505 A82505 Sequence  
c 29 11.8 59.0 41 9 A82526 A82526 Sequence  
c 30 11.8 59.0 41 9 AR064423 AR064423 Sequence  
c 31 11.8 59.0 41 9 AR083774 AR083774 Sequence  
c 32 11.8 59.0 41 10 I23386 I23386 Sequence  
c 33 11.8 59.0 41 10 I49918 I49918 Sequence  
c 34 11.8 59.0 41 10 I60381 I60381 Sequence  
c 35 11.8 59.0 41 10 I85612 I85612 Sequence  
c 36 11.8 59.0 49 97 S81487 S81487 T cell anti  
c 37 11.8 59.0 50 9 AR032874 AR032874 Sequence  
c 38 11.8 59.0 50 9 AR032875 AR032875 Sequence  
c 39 11.8 59.0 50 10 I29614 I29614 Sequence  
c 40 11.8 59.0 50 10 I29615 I29615 Sequence  
c 41 11.8 59.0 50 10 I91288 I91288 Sequence  
c 42 11.8 59.0 50 10 I91289 I91289 Sequence  
c 43 11.6 58.0 22 9 AR001375 AR001375 Sequence  
c 44 11.6 58.0 22 9 AR04862 AR04862 Sequence  
c 45 11.6 58.0 22 9 AR052248 AR052248 Sequence

## ALIGNMENTS

RESULT 1  
AR084719  
LOCUS AR084719 30 bp DNA  
DEFINITION Sequence 6 from patent US 5981215.  
ACCESSION AR084719  
VERSION AR084719.1 GI:10011489  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Weissner, P.S. and Coleman, T.A.  
TITLE Human cryptin growth factor  
JOURNAL Patent: US 5981215-A 6 09-NOV-1999;  
FEATURES Location/Qualifiers  
source 1..30  
BASE COUNT 11 a 7 c 3 g 9 t  
ORIGIN  
Query Match 69.0%; Score 13.8; DB 9; Length 30;  
Best Local Similarity 88.2%; Pred. No. 5e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 aagcagctgccattatt 20  
|| ||||| |||||  
Db 4 AACCACTGCTATTATT 20

RESULT 2

AR4595  
LOCUS AR4595 30 bp DNA  
DEFINITION Sequence 5 from Patent WO9845450.  
ACCESSION AR4595  
VERSION AR4595.1 GI:6733511  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Atkinson, E.M. and Kealey, J.T.  
TITLE PURIFIED TELOMERASE  
JOURNAL Patent: WO 9845450-A 5 15-OCT-1998;  
GERON CORP. (US)  
FEATURES Location/Qualifiers  
source 1..30  
/organism="unidentified"  
/db\_xref="taxon:32644"  
modified\_base 1  
BASE COUNT 6 a 6 c 7 g 10 t  
ORIGIN  
Query Match 68.0%; Score 13.6; DB 9; Length 30;  
Best Local Similarity 80.0%; Pred. No. 6.6e+03;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggtaagcagctgccattatt 20  
|| ||||| ||||| ||  
Db 6 GCCCAGCAGCTGACATTTT 25

RESULT 3  
AR079892  
LOCUS AR079892 30 bp DNA  
DEFINITION Sequence 5 from patent US 5968506.  
ACCESSION AR079892  
VERSION AR079892.1 GI:10006645  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Weinrich, S.L., Atkinson, E.M. III, Lichtsteiner, S.P., Vasserot, A.P.,  
Pruzan, R.A. and Kealey, J.T.  
TITLE Purified telomerase  
JOURNAL Patent: US 5968506-A 5 19-OCT-1999;  
FEATURES Location/Qualifiers  
source 1..30  
/organism="unknown"  
BASE COUNT 6 a 6 c 7 g 10 t  
ORIGIN  
Query Match 68.0%; Score 13.6; DB 9; Length 30;  
Best Local Similarity 80.0%; Pred. No. 6.6e+03;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggtaagcagctgccattatt 20  
|| ||||| ||||| ||  
Db 6 GCCCAGCAGCTGACATTTT 25

RESULT 4  
A97844/c  
LOCUS A97844 18 bp DNA  
DEFINITION Sequence 121 from Patent WO9914377.  
ACCESSION A97844  
VERSION A97844.1 GI:6781082  
KEYWORDS  
SOURCE unidentified.

ORGANISM unidentified  
unclassified.  
REFERENCE 1 (bases 1 to 18)  
AUTHORS Quint, W. and Kleter, B.  
TITLE DETECTION AND IDENTIFICATION OF HUMAN PAPILLOMAVIRUS BY PCR AND  
TYPE-SPECIFIC REVERSE HYBRIDIZATION  
JOURNAL Patent: WO 9914377-A 121 25-MAR-1999;  
INNOGENETICS NV (BE); DELFTS DIAGNOSTIC LAB B V (NL)  
FEATURES  
source 1..18  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"  
modified\_base 12  
/mod\_base=1  
BASE COUNT 5 a 2 c 5 g 5 t 1 others  
ORIGIN

Query Match 67.0%; Score 13.4; DB 9; Length 18;  
Best Local Similarity 87.5%; Pred. No. 8.4e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 agcagctgccattatt 20  
||||| |||||||||  
Db 16 AGCANATGCCATTATT 1

RESULT 5  
I44719/c I44719 18 bp DNA PAT 07-OCT-1997  
LOCUS  
DEFINITION Sequence 45 from patent US 5635384.  
ACCESSION I44719  
VERSION I44719.1 GI:2469432  
KEYWORDS  
SOURCE  
ORGANISM  
Unkown.  
Unclassified.  
REFERENCE 1 (bases 1 to 18)  
AUTHORS Walsh, T.A., Hey, T.D. and Morgan, A.E.R.  
TITLE Ribosome-inactivating proteins, inactive precursor forms thereof, a  
process for making and a method of using  
JOURNAL Patent: US 5635384-A 45 03-JUN-1997;  
FEATURES  
source 1..18  
Location/Qualifiers  
/organism="unknown"  
BASE COUNT 5 a 4 c 4 g 5 t  
ORIGIN

Query Match 67.0%; Score 13.4; DB 10; Length 18;  
Best Local Similarity 93.3%; Pred. No. 8.4e+03;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 taagcagctgccatt 17  
||||| |||||||||  
Db 17 TAAGCAGCTGCAATT 3

RESULT 6  
I52125/c I52125 18 bp DNA PAT 07-OCT-1997  
LOCUS  
DEFINITION Sequence 45 from patent US 5646026.  
ACCESSION I52125  
VERSION I52125.1 GI:2473326  
KEYWORDS  
SOURCE  
ORGANISM  
Unkown.  
Unclassified.  
REFERENCE 1 (bases 1 to 18)  
AUTHORS Walsh, T.A., Hey, T.D. and Morgan, A.E.R.  
TITLE Ribosome-inactivating proteins, inactive precursor forms thereof, a  
process for making and a method of using  
JOURNAL Patent: US 5646026-A 45 08-JUL-1997;

FEATURES  
source 1..18  
Location/Qualifiers  
/organism="unknown"  
BASE COUNT 5 a 4 c 4 g 5 t  
ORIGIN

Query Match 67.0%; Score 13.4; DB 10; Length 18;  
Best Local Similarity 93.3%; Pred. No. 8.4e+03;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 taagcagctgccatt 17  
||||| |||||||||  
Db 17 TAAGCAGCTGCAATT 3

RESULT 7  
AX096725/c  
LOCUS  
DEFINITION Sequence 1903 from Patent WO0118250.  
ACCESSION AX096725  
VERSION AX096725.1 GI:13512979  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
REFERENCE  
AUTHORS Lander, E.S., Gargill, M., Ireland, J.S., Bolk, S., Daley, G.O. and  
McCarthy, J.J.  
TITLE Single nucleotide polymorphisms in genes  
JOURNAL Patent: WO 0118250-A 1903 15-MAR-2001;  
WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH (US) ; Millennium  
Pharmaceuticals, Inc. (US)  
FEATURES  
source 1..21  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 8 a 3 c 4 g 5 t 1 others  
ORIGIN

Query Match 64.0%; Score 12.8; DB 10; Length 21;  
Best Local Similarity 77.8%; Pred. No. 1.9e+04;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gtaagcagctgccattat 19  
||||| :|||||||  
Db 18 GTAAAGTTRCTGCCATAAT 1

RESULT 8  
E04774/c  
LOCUS  
DEFINITION Synthetic DNA for site directed mutagenesis of protease derived  
from Bacillus NKS-21.  
ACCESSION E04774  
VERSION E04774.1 GI:2172970  
KEYWORDS JP 1993091876-A/13.  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS Takinishi, E., Kakinuma, S., Takemoto, A., Miyoda, Y. and Fukuyama, S.  
TITLE PROTEASE, ITS PRODUCTION AND USE  
JOURNAL Patent: JP 1993091876-A 13 16-APR-1993;  
SHOWA DENKO KK  
COMMENT  
OS Artificial gene  
OC Artificial sequence; Genes.  
OS Bacillus NKS-21  
PN JP 1993091876-A/13  
PD 16-APR-1993  
PF 02-OCT-1991 JP 1991280313

PI TAKINISHI EIKO, KAKINUMA SHINJI, TAKEMOTO AYANO, PI MIYODA  
YOSHIAKI,  
PI FUKUYAMA SHIRO  
PC C12N9/54.C11D3/386.C12N1/21.C12N15/57.(C12N9/54.C12R1:07), PC  
(C12N1/21.  
PC C12R1:07);  
CC strandedness: Single;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;

FEATURES  
source  
1..26  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 7 a 5 c 5 g 9 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 10; Length 26;  
Best Local Similarity 87.5%; Pred. No. 1.9e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 agcagctgcattatt 20  
||||| |||||  
DB 18 AGCAGCTGCAATTAAT 3

RESULT 9  
I06269 106269 32 bp PAT 02-DEC-1994  
LOCUS Sequence 27 from Patent EP 0319052.  
DEFINITION I06269  
ACCESSION I06269  
VERSION I06269.1 GI:590260  
KEYWORDS  
SOURCE Unknown.  
ORGANISM  
Unclassified.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Thomas Jnr,K.A. and Linemeyer,D.L.  
TITLE Mutant acidic fibroblast growth factor  
JOURNAL Patent: EP 0319052-A2 27 07-JUN-1989;  
FEATURES Location/Qualifiers  
source  
1..32  
/organism="unknown"

BASE COUNT 5 a 8 c 8 g 11 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 32;  
Best Local Similarity 78.9%; Pred. No. 2.6e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gtaagcagctgcattatt 20  
||| ||||| |||||  
DB 7 GTTTACAGCTGCCATTTT 25

RESULT 10  
AX094845/c  
LOCUS AX094845 21 bp DNA PAT 30-MAR-2001  
DEFINITION Sequence 23 from Patent WO0118250.  
ACCESSION AX094845  
VERSION AX094845.1 GI:13511048  
KEYWORDS  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Lander,E.S., Gargill,M., Ireland,J.S., Bolk,S., Daley,G.Q. and  
McCarthy,J.J.  
TITLE Single nucleotide polymorphisms in genes  
JOURNAL Patent: WO 0118250-A 23 15-MAR-2001;

WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH (US) ; Millennium  
Pharmaceuticals, Inc. (US)  
FEATURES Location/Qualifiers  
source  
1..21  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 6 a 6 c 4 g 4 t 1 others  
ORIGIN

Query Match 62.0%; Score 12.4; DB 10; Length 21;  
Best Local Similarity 81.2%; Pred. No. 3.3e+04;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 gtaagcagctgcatt 17  
||| ||||| |||||  
DB 18 GTAGGCARCTGCGATT 3

RESULT 11  
AX095462/c  
LOCUS AX095462 21 bp DNA PAT 30-MAR-2001  
DEFINITION Sequence 640 from Patent WO0118250.  
ACCESSION AX095462  
VERSION AX095462.1 GI:13511665  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 21)  
AUTHORS Lander,E.S., Gargill,M., Ireland,J.S., Bolk,S., Daley,G.Q. and  
McCarthy,J.J.  
TITLE Single nucleotide polymorphisms in genes  
JOURNAL Patent: WO 0118250-A 640 15-MAR-2001;  
WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH (US) ; Millennium  
Pharmaceuticals, Inc. (US)  
FEATURES Location/Qualifiers  
source  
1..21  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 7 a 2 c 6 g 5 t 1 others  
ORIGIN

Query Match 62.0%; Score 12.4; DB 10; Length 21;  
Best Local Similarity 81.2%; Pred. No. 3.3e+04;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 agcagctgcattatt 20  
||| ||||| |||||  
DB 21 AACAGCGCGCCTTATT 6

RESULT 12  
AR032341/c  
LOCUS AR032341 39 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 24 from patent US 5869055.  
ACCESSION AR032341  
VERSION AR032341.1 GI:5947946  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 39)  
AUTHORS Juan,S., Lichenstein,H.S. and Wright,S.D.  
TITLE Anti-inflammatory CD14 polypeptides  
JOURNAL Patent: US 5869055-A 24 09-FEB-1999;  
FEATURES Location/Qualifiers  
source  
1..39  
/organism="unknown"

BASE COUNT 7 a 13 c 13 g 6 t  
ORIGIN

11

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:38 ; Search time 876.95 Seconds  
(without alignments)  
14.320 Million cell updates/sec

Title: US-09-757-100b-10

Perfect score: 20

Sequence: 1 ggtaagcagctgccattatt 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	20	100.0	20	22	AAC65542 Human focal adhesi
c 2	16	80.0	22	20	AA336327 Sense primer used
c 3	15-2	76.0	29	16	AAT22589 Human gene signatu
4	15	75.0	15	22	AAC65562 Human focal adhesi
c 5	14.4	72.0	49	21	AAA64261 PCR primer for hum
6	14	70.0	31	20	AA06286 Human biallelic po
7	13.8	69.0	30	18	AAT51062 Human criptin grow
8	13.8	69.0	40	20	AA86412 PCR primer PDZK5.5
9	13.6	68.0	30	19	AAV63648 Antisense oligonuc
10	13.6	68.0	30	20	AAT23630 Human clone 28-1 t
c 11	13.2	66.0	21	22	AAF97138 Human gene single

12	13.2	66.0	29	21	AAA03899 Polymorphic fragme
c 13	13.2	66.0	30	16	AA086352 Mutagenic oligo fo
14	13.2	66.0	48	21	AAAL3301 PCR primer NotI.E5
c 15	13.2	66.0	48	21	AAAL3302 PCR primer ES.NotI
16	12.8	64.0	20	20	AAZ01868 PCR primer used to
c 17	12.8	64.0	21	22	AAF95258 Human gene single
c 18	12.6	63.0	21	22	AAF95875 Human gene single
c 19	12.6	63.0	29	21	AAZ60352 PCR primer RP-6A u
c 20	12.6	63.0	44	18	AA778806 Kappa light chain
c 21	12.6	63.0	44	19	AAV39247 Primer o-551 used
c 22	12.6	63.0	44	20	AAZ22001 Oligonucleotide us
c 23	12.6	63.0	49	21	AAZ60353 PCR primer RP-6B u
c 24	12.4	62.0	27	20	AAZ01016 PCR primer for PGI
25	12.4	62.0	31	20	AAZ06130 Human biallelic po
c 26	12.2	61.0	24	21	AAZ74420 Human c-myc cDNA P
27	12.2	61.0	25	17	AAT34140 Primer for icosape
28	12.2	61.0	25	17	AAT10302 RNA component of h
29	12.2	61.0	25	19	AAV00513 Primer for S. putr
30	12.2	61.0	25	20	AAZ07280 Human telomerase R
31	12.2	61.0	27	19	AAV21898 Nuclease resistant
32	12.2	61.0	27	19	AAV21898 Nuclease resistant
c 33	12.2	61.0	30	21	AAZ18958 Human Meg-3 DNA pr
c 34	12.2	61.0	32	19	AAV37859 PEA toxin gene and
c 35	12.2	61.0	32	19	AAV37851 Transformed lympho
c 36	12.2	61.0	40	18	AAZ57948 L-selectin truncat
c 37	12.2	61.0	45	21	AAZ88569 BYDV-PAV Illinois
c 38	12.2	61.0	49	18	AAZ57947 L-selectin truncat
c 39	12.2	61.0	49	20	AAZ86109 Nucleic acid ligand
c 40	12	60.0	29	18	AAZ96994 Human GDP-fucose p
c 41	12	60.0	29	18	AAZ97584 Human GDP-fucose p
c 42	12	60.0	39	19	AAV29643 Nucleotide sequenc
c 43	12	60.0	40	19	AAV28321 24 gene cap site,
c 44	12	60.0	42	16	AAZ04808 B.stearothermophil
c 45	12	60.0	44	13	AAZ25193 TAP probe (1). Sy

ALIGNMENTS

RESULT 1  
AAC65542  
ID AAC65542 standard; DNA; 20 BP.  
XX  
AC AAC65542;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human focal adhesion kinase antisense sequence #8.  
XX  
KW Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.  
XX  
OS Homo sapiens.  
XX  
PN US6133031-A.  
XX  
PD 17-OCT-2000.  
XX  
PF 19-AUG-1999; 99US-0377310.  
XX  
PR 19-AUG-1999; 99US-0377310.  
XX  
PA (ISIS-) ISIS PHARM INC.  
XX  
PI Monia BP, Gaarde WA;  
XX  
DR WPI; 2001-006141/01.  
XX  
PT New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
XX neovascularization, or for diagnosing and treating e.g. colon cancer -

PS Example 2; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, embryonic  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

XX Sequence 20 BP; 5 A; 4 C; 5 G; 6 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.29; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtaagcagctgccattatt 20  
DB 1 ggtaagcagctgccattatt 20

RESULT 2

AAX36327/C  
ID AAX36327 standard; DNA; 22 BP.

XX

AC AAX36327;

DT 21-JUL-1999 (first entry)

DE Sense primer used to amplify human FAK cDNA.

XX Human; tumorigenicity; glycosyltransferase;  
KW malignancy; brain cancer cell; protein glycosylation; glioma;  
KW meningioma; brain tumour; FAK; PCR primer; ss.

XX Synthetic.

XX WO924584-A1.

XX 20-MAY-1999.

XX 12-NOV-1998; 98WO-US24224.

XX 12-NOV-1997; 97US-0969437.

XX (NEUR-) NEUROTHERAPEUTICS.

XX Moskal JR, Yamamoto H;

XX WPI; 1999-327411/27.

XX Altering tumorigenicity and malignancy of brain cancer cells

XX Example 3; Page 34; 83pp; English.

XX PCR primers AAX36327-28 were used to amplify human FAK cDNA, in the  
CC course of the invention. The specification describes a method for  
CC altering the tumorigenicity or malignancy of brain cancer cells by  
CC changing the activity of glycosyltransferase in the cell so that  
CC glycosylation of cellular proteins is modified. The method is applied  
CC to glioma or meningioma, for prevention or treatment of brain tumours.  
CC Measuring the level of glycosyltransferase expression in brain cells  
CC is used to detect or predict their tumorigenicity.

XX Sequence 22 BP; 4 A; 7 C; 5 G; 6 T; 0 other;

Query Match 80.0%; Score 16; DB 20; Length 22;

Best Local Similarity 100.0%; Pred. No. 33; Mismatches 16; Conservative 0; Indels 0; Gaps 0;

QY 1 ggtaagcagctgccatt 16

DB 16 GGTAAAGCAGCTGCCAT 1  
|||||

RESULT 3

AAT22589/C

ID AAT22589 standard; cDNA to mRNA; 29 BP.

XX

AC AAT22589;

DT 01-OCT-1996 (first entry)

DE Human gene signature HUMGS04205.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; Cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.

XX Homo sapiens.

XX WO9514772-A1.

XX 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP01916.

XX 12-NOV-1993; 93JP-0355504.

XX (MATS/) MATSUBARA K.

XX (OKUB/) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI; 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

XX Claim 1; Page 1163; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in AAT19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared from the  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

XX Sequence 29 BP; 12 A; 6 C; 6 G; 5 T; 0 other;

Query Match 76.0%; Score 15.2; DB 16; Length 29;

Best Local Similarity 85.0%; Pred. No. 88;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtaagcagctgccattatt 20

DB 26 GGTACAGCAGCTACCTTTATT 7

RESULT 4

AAC65562

ID AAC65562 standard; DNA; 15 BP.



```
XX AAC65562;
XX
XX 12-FEB-2001 (first entry)
DT
XX
XX Human focal adhesion kinase antisense sequence #28.
DE
XX
XX Human; focal adhesion kinase; FAK; signal transduction; cancer;
KW embryonic development disorder; angiogenic disorder; wound healing;
KW antisense; phosphorothioate; ss.
XX
XX Homo sapiens.
OS
XX
XX US6133031-A.
PN
XX
XX 17-OCT-2000.
PD
XX
XX 19-AUG-1999; 99US-0377310.
PF
XX
XX 19-AUG-1999; 99US-0377310.
PR
XX
XX (ISIS-) ISIS PHARM INC.
PA
XX
XX Monia BP, Gaarde WA;
XX
XX WPI; 2001-006141/01.
DR
XX
XX New antisense compounds for inhibiting focal adhesion kinase
PT expression, especially useful for inhibiting retinal
PT neovascularization, or for diagnosing and treating e.g. colon cancer -
XX
XX Claim 15; Column 25; 30pp; English.
XX
XX The present invention describes a number of phosphorothioate antisense
CC sequences to the human focal adhesion kinase (FAK) protein. This protein
CC is involved in integrin-mediated signal transduction, and is implicated
CC in cancer, particularly colon, breast and oral tumours, embryonic
CC development disorders, angiogenic disorders and wound healing. The
CC antisense sequences, including the one shown here, can be used in the
CC treatment of all of these.
XX
XX Sequence 15 BP; 4 A; 4 C; 3 G; 4 T; 0 other;
SQ

Query Match 75.0%; Score 15; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 taagcagctgccatt 17
Db 1 taagcagctgccatt 15
|||||

RESULT 5
AAA64261/c
ID AAA64261 standard; DNA; 49 BP.
XX
XX
XX AAA64261;
AC
XX
XX 20-DEC-2000 (first entry)
DT
XX
XX PCR primer for human fibronectin collagen-binding domain cDNA.
DE
XX
XX Fibronectin; collagen-binding domain; sustained release; gene therapy;
KW physiologically active polypeptide; topical retention; PCR primer;
KW tissue regeneration; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200049159-A1.
PN
XX
XX 24-AUG-2000.
PD
XX
XX
```

```
PF 21-FEB-2000; 2000WO-JP00964.
XX
XX 19-FEB-1999; 99JP-0041913.
PR 01-NOV-1999; 99JP-0311364.
XX
XX (TERU ) TERUMO CORP.
PA
XX
XX Ishikawa T, Kitajima T;
PI
XX
XX WPI; 2000-565375/52.
DR
XX
XX Collagen-binding active polypeptide for use in an agent for enabling
PT topical retention or sustained release of a physiologically active
PT peptide or physiological activity-imparting agent comprises a
PT fibronectin peptide -
XX
XX Example 1; Page 124; 135pp; English.
PS
XX
XX PCR primers AAA64261-62 were used to amplify cDNA encoding a human
CC fibronectin collagen-binding domain. The amplified sequence was
CC used to construct a collagen-binding physiologically active polypeptide.
CC This polypeptide comprises a peptide from fibronectin ligated to a
CC physiologically active peptide. The polypeptides are used in an agent
CC for enabling topical retention or sustained release of a physiologically
CC active peptide or physiological activity-imparting agent. They may
CC be used in gene therapy and in tissue regeneration.
XX
XX Sequence 49 BP; 12 A; 16 C; 11 G; 10 T; 0 other;
SQ

Query Match 72.0%; Score 14.4; DB 21; Length 49;
Best Local Similarity 93.8%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggtaagcagctgccatt 16
Db 33 GGTAACAGCTGCCAT 18
|||||

RESULT 6
AAx06286
ID AAX06286 standard; DNA; 31 BP.
XX
XX
XX AAX06286;
AC
XX
XX 31-MAR-1999 (first entry)
DT
XX
XX Human biallelic polymorphic DNA fragment SGC30775.
DE
XX
XX Polymorphism; biallelic; paternity testing; forensic; genetic mapping;
KW phenotypic typing; medicament; disease; marker; human; ss.
XX
XX Homo sapiens.
OS
XX
XX WO9858529-A2.
PN
XX
XX 30-DEC-1998.
PD
XX
XX 22-JUN-1998; 98WO-US12930.
PF
XX
XX 24-JUN-1997; 97US-0050594.
PR
XX
XX (AFFY-) AFFYMETRIX INC.
PA
XX
XX Berno A, Chee M, Fan J, Lipshutz RJ;
PI
XX
XX WPI; 1999-080963/07.
DR
XX
XX New nucleic acid segments containing polymorphic sites - used for,
PT e.g. detecting a disease phenotype, in forensics, paternity testing
PT or genetic mapping of phenotypic traits
XX
XX Claim 1; Page 18; 61pp; English.
PS
```

xx Sequences AAX06101-X06558 represent human DNA fragments which contain  
 CC biallelic polymorphic markers. The base occupying the polymorphic site  
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These  
 CC fragments can be used in a method for determining polymorphic forms in  
 CC an individual. The invention further provides computer-readable storage  
 CC medium for storing data for access by an application programme being  
 CC executed on a data processing system. Such a method comprises a data  
 CC structure stored in the computer-readable storage medium, the data  
 CC structure including information resident in a database used by the  
 CC application programme and including records, each record comprising  
 CC information identifying a polymorphism shown in the above sequences. The  
 CC products and methods can be used for analysing polymorphic sites in  
 CC individuals for testing for the presence of a disease phenotype or in  
 CC forensics, paternity testing or genetic mapping of phenotypic traits.  
 CC They can also be used for the production of polypeptides expressed by  
 CC variant genes and for the production of transgenic animals. The nucleic  
 CC acid segments can also be used in the manufacture of medicaments for the  
 CC treatment or prophylaxis of diseases.  
 XX  
 SQ Sequence 31 BP; 5 A; 7 C; 6 G; 12 T; 1 other;

Query Match 70.0%; Score 14; DB 20; Length 31;  
 Best Local Similarity 87.5%; Pred. No. 3.7e+02;  
 Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 agcagctgccattatt 20  
 ||||| ||||| |||||  
 Db 6 agcagctgccrttatt 21

RESULT 7  
 AAT51062  
 ID AAT51062 standard; cDNA; 30 BP.

XX  
 XX AAT51062;  
 DT 05-APR-1997 (first entry)  
 DE Human criptin growth factor 3' primer for COS expression.

XX Criptin growth factor; CGF; angiogenesis; wound healing; vulnery;  
 KW muscle wastage; osteoporosis; implant fixation; tissue regeneration;  
 KW pancreas cancer; diagnosis; gene therapy; primer; PCR;  
 KW polymerase chain reaction; COS; ss.  
 XX  
 OS Synthetic.

XX WO9639420-A1.

XX 12-DEC-1996.

XX 05-JUN-1995; 95WO-US07087.

XX 05-JUN-1995; 95WO-US07087.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Coleman TA, Meissner PS;

XX WPI; 1997-043055/04.

XX New isolated human Criptin Growth Factor polypeptide - which can be  
 PT used to stimulate angiogenesis and develop products for use in  
 PT diagnosis and therapy

XX Example 3; Page 36; 52pp; English.

XX A 3' PCR primer (AAT51062) contains a PvuII site followed by 15  
 CC nucleotides that are the reverse complement of human criptin  
 CC growth factor (CGF) 3' DNA starting at the translational stop codon  
 CC (see also AAT51058). It was used with a 5' primer (AAT51061) to

CC amplify the CGF DNA clone deposited as ATCC 97142. The PCR product  
 CC was incorporated into plasmid pN346 to allow prodn. of CGF (see  
 CC also AAW09111) in transfected COS cells.

XX Sequence 30 BP; 11 A; 7 C; 3 G; 9 T; 0 other;

Query Match 69.0%; Score 13.8; DB 18; Length 30;  
 Best Local Similarity 88.2%; Pred. No. 4.6e+02;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 aagcagctgccattatt 20  
 || ||||| |||||  
 Db 4 aaccagctgctattatt 20

RESULT 8  
 AAX86412  
 ID AAX86412 standard; DNA; 40 BP.

XX AC AAX86412;

XX DT 29-SEP-1999 (first entry)

XX PCR primer PDZK5.5C used to amplify DNA encoding MMS1 protein.

XX Human; MMS1 protein; MMAC1 interacting protein; tumour suppression;  
 KW MMAC1 pathway; immunogen; cancer; cell neoplastic growth; PCR primer; ss.

XX Synthetic.  
 OS Homo sapiens.

XX WO9936566-A1.

XX PD 22-JUL-1999.

XX PF 19-JAN-1999; 99WO-US00995.

XX PR 20-JAN-1998; 98US-0071861.

XX (MYRI-) MYRIAD GENETICS INC.

XX Bartel PL, Tavtigian SV;

XX WPI; 1999-458472/38.

XX MMS1, an MMAC1 (tumour suppressor) interacting protein and related  
 PT polynucleotides

XX Example 5; Page 51; 107pp; English.

XX PCR primers AAX86368-X86423 were used to amplify DNA encoding a human  
 CC MMS1 protein. The PCR templates were derived from tumour cell lines,  
 CC and the amplicons were tested for mutations. The MMS1 protein is a  
 CC MMAC1 interacting protein which is involved in tumour suppression  
 CC activity in the MMAC1 pathway. MMS1, antigenic fragments or fusion  
 CC proteins of these are used as immunogens for antibody production. Primers  
 CC derived from MMS1 genomic clones can be used for identification of MMS1  
 CC genes and for synthesis by amplification of MMS1 DNA or RNA. Detecting  
 CC an alteration in MMS1 can be used to diagnose cancer. A germline  
 CC alteration in an MMS1 gene is indicative of a predisposition to cancer.  
 CC A somatic mutation in an MMS1 gene is indicative that the tissue is  
 CC cancerous. Analysis of MMAC1 and MMS1 (or PDZ domain 6 of MMS1)  
 CC binding interactions can be used for detection of alterations in MMAC1  
 CC associated with cancer. Wild-type MMS1 or a homologue can be used to  
 CC supply wild-type MMS1 gene function (or a substantially similar  
 CC function) to a cell, which has lost the gene function due to a MMS1  
 CC gene mutation. The gene suppresses neoplastic growth of the cell.  
 CC transgenic animals having an altered MMS1 can be used as a model for  
 CC identifying drug candidates useful in treating cancer.

XX Sequence 40 BP; 9 A; 12 C; 7 G; 12 T; 0 other;

Query Match 69.0%; Score 13.8; DB 20; Length 40;  
 Best Local Similarity 88.2%; Pred. No. 4.8e+02;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 aagcagctgccattatt 20  
 ||| ||||| |||||  
 Db 16 acgcatcgccattatt 32

RESULT 9  
 AAV63648  
 ID AAV63648 standard; DNA; 30 BP.  
 AC AAV63648;  
 XX  
 DT 15-FEB-1999 (first entry)  
 XX  
 DE Antisense oligonucleotide 13 for human telomerase RNA component.  
 XX  
 KW Human; telomerase RNA component; anticancer therapy; purification;  
 KW assay; vaccine; cancer; antisense oligonucleotide; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT modified\_base 1 /\*tag= a  
 FT /note= "biotinylated"

XX WO9845450-A1.  
 XX  
 XX 15-OCT-1998.  
 XX  
 PF 04-APR-1997; 97WO-US06012.  
 XX  
 PR 04-APR-1997; 97WO-US06012.  
 XX  
 PA (GERO-) GERON CORP.  
 XX

PI Atkinson EM, Kealey JT, Lichtsteiner SP, Pruzan RA;  
 PI Vasserot AP, Weinrich SL;  
 XX WPI; 1998-594485/50.  
 XX  
 PT Purification of telomerase on affinity material - useful for, e.g.  
 PT diagnosis and treatment of cancer  
 XX  
 PS Disclosure; Page 24; 76pp; English.

XX The present sequence represents an antisense oligonucleotide  
 CC directed against the human telomerase RNA component gene sequences.  
 CC The oligonucleotide can be used as an affinity agent in the methods of  
 CC the invention, which are used to purify human telomerase. The methods  
 CC involve the use of several sequential steps, including the use of two  
 CC matrices that bind molecules bearing negative charges, a matrix that  
 CC binds molecules bearing positive charges, an affinity purification step  
 CC and a size separation. Telomerase is a particular target of anticancer  
 CC therapies, and is useful in assays for characterizing (pre)cancerous  
 CC cells. Telomerase can also be used to screen for specific modulators,  
 CC for biochemical analysis of its activity, and in preparation of  
 CC antibodies. Fragments of telomerase, or nucleic acid encoding them,  
 CC are used in vaccines, and for treating over expression of telomerase,  
 CC particularly in cancer.

XX Sequence 30 BP; 6 A; 6 C; 8 G; 10 T; 0 other;

Query Match 68.0%; Score 13.6; DB 19; Length 30;  
 Best Local Similarity 80.0%; Pred. No. 5.8e+02;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggtaagcagctgccattatt 20  
 ||| ||||| ||||| ||  
 Db 6 ggccagcagctgacatttt 25

RESULT 10  
 AAZ23630  
 ID AAZ23630 standard; DNA; 30 BP.  
 XX  
 AC AAZ23630;  
 XX  
 DT 07-JAN-2000 (first entry)  
 XX  
 DE Human clone 28-1 telomerase oligonucleotide oligo-13.

XX Telomerase; human; immune response; cancer; vaccine; treatment;  
 KW disease; primer; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT modified\_base 1 /\*tag= a  
 FT /note= "5'-biotinylated guanosine"

XX US5968506-A.  
 XX  
 XX 19-OCT-1999.  
 PD  
 XX 04-APR-1997; 97US-0833377.  
 XX  
 XX 04-AUG-1995; 95US-0510736.  
 XX  
 XX (GERO-) GERON CORP.

XX Atkinson EM, Lichtsteiner SP, Weinrich SL, Pruzan RA, Kealey JT;  
 XX Vasserot AP;  
 XX WPI; 1999-590379/50.  
 XX  
 PT Compositions comprising human telomerase, useful for treating diseases  
 PT associated with overexpression of telomerase e.g. cancer -  
 XX  
 PS Disclosure; Column 45-46; 34pp; English.

XX This invention describes a novel composition comprising human telomerase  
 CC having at least 2000-fold (preferably at least 6000-fold) increased  
 CC relative purity compared with crude extract of cells from  
 CC adenovirus-transformed kidney cell line. The composition is useful for  
 CC eliciting an immune response in animals and may therefore be used as a  
 CC vaccine for treating diseases associated with the overexpression of  
 CC telomerase e.g. cancer. AAZ23626-223637 represent oligonucleotides used  
 CC in the isolation of human clone 28-1 which contains a fragment of the  
 CC human telomerase described in the method of the invention.

XX Sequence 30 BP; 6 A; 6 C; 8 G; 10 T; 0 other;

Query Match 68.0%; Score 13.6; DB 20; Length 30;  
 Best Local Similarity 80.0%; Pred. No. 5.8e+02;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggtaagcagctgccattatt 20  
 ||| ||||| ||||| ||  
 Db 6 ggccagcagctgacatttt 25

RESULT 11  
 AAF97138/c  
 ID AAF97138 standard; DNA; 21 BP.  
 XX  
 AC AAF97138;

```

XX 06-JUN-2001 (first entry)
XX Human gene single nucleotide polymorphism #1899.
DE Human; variant thrombospondin 1; variant thrombospondin 4; SNP;
KW polymorphism; vascular disease; coronary artery disease; forensics;
KW myocardial infarction; atherosclerosis; stroke; venous thromboembolism;
KW pulmonary embolism; paternity test; ds.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Variation replace(11,T)
FT /*tag= a
FT /standard_name= "single nucleotide polymorphism"
XX
PN WO200118250-A2.
XX
PD 15-MAR-2001.
XX
PF 07-SEP-2000; 2000WO-US24503.
XX
PR 10-SEP-1999; 99US-0153357.
PR 26-JUL-2000; 2000US-0220947.
PR 16-AUG-2000; 2000US-0225724.
XX
PA (WHEED ) WHITEHEAD INST BIOMEDICAL RES.
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Lander ES, Gargill M, Ireland JS, Boik S, Daley GO, McCarthy JJ;
XX WPI; 2001-226749/23.
XX
XX Nucleic acids comprising single nucleotide polymorphisms useful in
PT applications such as forensics, paternity testing, medicine, genetic
PT analysis and phenotype correlations to diseases such as diabetes and
PT atherosclerosis .
XX
PS Examples; Page 177; 242pp; English.
XX
CC The present invention provides a method of diagnosing a vascular disease
CC in an individual, involving determining the sequence at various
CC polymorphic sites within the human thrombospondin 1 and thrombospondin 4
CC genes. The sequences at a number of polymorphic sites are also provided
CC in the specification. In particular, the method can be used in the
CC diagnosis of atherosclerosis, myocardial infarction, coronary heart
CC disease, stroke, peripheral vascular diseases, venous thromboembolism
CC and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also
CC useful in forensics, paternity testing, genetic analysis and phenotype
CC correlations to diseases. The present sequence is an example of one of
CC the human gene SNPs shown in the specification.
XX
SQ Sequence 21 BP; 8 A; 4 C; 4 G; 5 T; 0 Other;

```

```

Query Match 66.0%; Score 13.2; DB 22; Length 21;
Best Local Similarity 83.3%; Pred. NO. 8.9e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 2 gtaagcagctgcattat 19
    ||||| ||||| |||
Db 18 GTAAGTTCCTGCCATAAT 1

```

```

RESULT 12
AA03899
ID AAA03899 standard; DNA; 29 BP.
XX
XX AAA03899;
AC
XX
DT 22-MAY-2000 (first entry)
XX

```

```

DE Polymorphic fragment of hypertension associated gene AGT.
XX
XX Polymorphism; hypertension; agammaglobulinemia; diabetes insipidus;
KW Leach-Nyhan syndrome; muscular dystrophy; Wiskott-Aldrich syndrome;
KW Fabrys disease; familial hypercholesterolemia; hereditary spherocytosis;
KW polycystic kidney disease; von Willebrands disease; hereditary telangiectasia;
KW tuberous sclerosis; hereditary hemorrhagica telangiectasia;
KW familial colonic polyposis; osteogenesis imperfecta; porphyria;
XX Ehlers-Danlos syndrome; ss.
XX Homo sapiens.
XX
XX EP955382-A2.
XX
XX 10-NOV-1999.
XX
XX 07-MAY-1999; 99EP-0250150.
XX
XX 07-MAY-1999; 98US-0084641.
PR 03-MAY-1999; 99US-0304232.
XX
XX (AFFY-) AFFYMETRIX INC.
XX (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX Fan JB, Chakravarti A, Haluska MK;
XX WPI; 2000-107928/10.
XX
XX Novel nucleic acids containing polymorphisms used in the diagnosis of
XX hypertension .
XX
XX Claim 1; Page 19; 53pp; English.
XX
XX The invention provides polymorphic fragments of genes associated with
XX hypertension. The nucleic acids including the polymorphic sites can be
XX used as probes or primers for expressing variant proteins. Detection of
XX the polymorphisms is useful in designing prophylactic and therapeutic
XX regimes customized to underlying abnormalities. The polymorphisms can be
XX used for association studies for hypertension, and in hypertension
XX diagnostic assays. Where the polymorphisms have strong correlation with
XX hypertension, within a gene, they are likely to have a causative role in
XX hypertension. This information can be used to find the precise role of a
XX polymorphism in the disease, and this can be used to identify potential
XX drugs which combat the disease. The polymorphisms can be tested for
XX association with other diseases e.g. agammaglobulinemia, diabetes
XX insipidus, Leach-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich
XX syndrome, Fabrys disease, familial hypercholesterolemia, polycystic
XX kidney disease, hereditary spherocytosis, von Willebrands disease,
XX tuberous sclerosis, hereditary hemorrhagica telangiectasia, familial
XX colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, and
XX acute intermittent porphyria. The polymorphic forms can also be used in
XX forensics to identify individuals.
XX
XX Sequence 29 BP; 5 A; 4 C; 12 G; 7 T; 1 Other;

```

```

Query Match 66.0%; Score 13.2; DB 21; Length 29;
Best Local Similarity 75.0%; Pred. NO. 9.3e+02;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 ggtaagcagctgcattatt 20
    ||| ||| ||||| |||
Db 2 gggagaagctgcygtgtt 21

```

```

RESULT 13
AAQ86352/c
ID AAQ86352 standard; DNA; 30 BP.
XX
XX AAQ86352;
AC
XX
DT 29-DEC-1995 (first entry)
XX

```

DE Mutagenic oligo for human protein S variant I425A/I426A.  
XX Protein S; PS; vitamin K-dependent protein; mutagenic oligo; ss.  
XX Synthetic.  
XX Key Location/Qualifiers  
FH misc\_difference 13..14 /\*tag= a  
FT /\*note= "nts for introducing desired AA substn."  
FT misc\_difference 16..17 /\*tag= b  
FT /\*note= "nts for introducing desired AA substn."  
XX US5405946-A.  
XX 11-APR-1995.  
XX 02-DEC-1992; 92US-0985691.  
XX 02-DEC-1992; 92US-0985691.  
XX (SCRI ) SCRIPPS RES INST.  
XX Bertina R, Bouma BN, Griffin JH;  
XX WPI; 1995-154630/20.  
XX  
XX New recombinant protein S variants - having reduced C4b binding  
PT protein binding activity and anticoagulant activity for treating  
PT thrombosis  
XX  
XX Example; Column 18; 24pp; English.  
XX  
XX For preparing a protein S (PS) expression vector, partial cDNAs  
CC coding for human protein S were first isolated as described by Ploos  
CC van Amstel et al., FEBS Lett., 222:186-190 (1987) from a pUC9 human  
CC liver cDNA library. The cDNA sequence is given in Q86348. The PS nt  
CC sequence is also listed in Genbank having the accession number Y00692.  
CC The mRNA encodes a preprotein having 676 AAs. After post-  
CC translation processing the corresp. translated mature PS consists of  
CC 635 AAs as given in R72350. The AA sequence is also listed in  
CC Genbank having the accession no. A26157. PS can be modified without  
CC significant loss of anticoagulant activity by introducing one or  
CC more mutations in the region between residues 425 and 432 to reduce  
CC significantly or eliminate the ability of PS to bind C4BP. A  
CC variant protein S is claimed having AA residue substitutions  
CC K429S, I425A/I426A and K432E (AA residue nos. corresp. to R72350).  
CC Oligos Q86349-Q86350 were used to produce mutations  
CC in the first disulfide loop in the sex hormone binding globulin-like  
CC domain of PS.  
XX  
XX Sequence 30 BP; 16 A; 4 C; 7 G; 3 T; 0 Other;

Query Match 66.0%; Score 13.2; DB 16; Length 30;  
Best Local Similarity 83.3%; Pred. No. 9.3e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 taagcagctgcccattatt 20  
| | | | | | | | | | | | | |  
Db 20 TGAGACGCTTCCTTTATT 3

RESULT 14  
AAAL3301  
ID AAAL3301 standard; cDNA; 48 BP.  
XX AC  
XX AAA13301;  
XX  
XX 24-JUL-2000 (first entry)  
XX

PCR primer NotI.E5 used in the production of vectors for trans-splicing.

XX Recombinant nucleic acid production; combinatorial gene library;  
KW ordered gene assembly; trans-splicing; tissue plasminogen activator;  
KW vector production; PCR primer; ss.  
XX Synthetic.  
XX WO200017342-A2.  
XX 30-MAR-2000.  
XX 21-SEP-1999; 99WO-US21929.  
XX 21-SEP-1998; 98US-0101328.  
XX 20-SEP-1999; 99US-0101328.  
XX (UYBO-) UNIV BOSTON.  
XX Jarrell KA, Mulcheeva S, Donahue W;  
XX WPI; 2000-303208/26.  
XX  
XX In vivo production of nucleic acid, useful e.g. for producing  
PT combinatorial gene libraries or ribozymes, by trans-splicing two RNAs  
PT containing exon and intron component  
XX  
XX Example 4; Page 49; 186pp; English.

XX This sequence represents a PCR primer used in the construction of  
CC trans-splicing vectors to illustrate the invention. The present invention  
CC relates to the in vivo production of recombinant nucleic acid sequences.  
CC The method comprises expressing in a cell, two transcripts, one  
CC containing a first exon and first intron component, and a second  
CC transcript comprising a second intron component and a second exon.  
CC Transcript 1 and transcript 2 are allowed to trans-splice, forming a  
CC product containing exon 1 and exon 2, but not the intron components. The  
CC invention makes use of the ability of intronic sequences derived from  
CC group I or group II introns to mediate specific cleavage and ligation of  
CC discontinuous nucleic acid molecules. The method is used to produce  
CC recombinant nucleic acids, their products, or ribozymes in vivo. The  
CC method is preferably used for the preparation of combinatorial gene  
CC libraries in which the order and composition of exons are random. The  
CC method is also used for ordered gene assembly, or for the assembly of  
CC genes ordered at some exons but randomized at others. New genes can be  
CC selected rapidly and efficiently, and a very wide range of exons may be  
CC trapped.

XX Sequence 48 BP; 16 A; 12 C; 10 G; 10 T; 0 other;

Query Match 66.0%; Score 13.2; DB 21; Length 48;  
Best Local Similarity 83.3%; Pred. No. 1e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gtaagcagctgcccattat 19  
| | | | | | | | | | | | | |  
Db 17 gaaagcgccgcccattat 34

RESULT 15  
AAAL3302/C  
ID AAAL3302 standard; cDNA; 48 BP.  
XX  
XX AAA13302;  
XX

XX 24-JUL-2000 (first entry)

XX PCR primer E5.NotI used in the production of vectors for trans-splicing.  
KW Recombinant nucleic acid production; combinatorial gene library;  
KW ordered gene assembly; trans-splicing; tissue plasminogen activator;  
KW vector production; PCR primer; ss.  
XX

OS Synthetic.  
XX WO200017342-A2.  
XX  
XX  
PD 30-MAR-2000.  
XX  
XX 21-SEP-1999; 99WO-US21929.  
XX  
XX 21-SEP-1998; 98US-0101328.  
PR 20-SEP-1999; 99US-0101328.  
XX  
XX (UYBO-) UNIV BOSTON.  
PA  
XX Jarrell KA, Mulcheeva S, Donahue W;  
PI WPI; 2000-303208/26.  
XX  
DR  
XX  
XX In vivo production of nucleic acid, useful e.g. for producing  
PT combinatorial gene libraries or ribozymes, by trans-splicing two RNAs  
PT containing exon and intron component  
XX  
XX Example 4; Page 49; 186pp; English.  
PS  
XX This sequence represents a PCR primer used in the construction of  
CC trans-splicing vectors to illustrate the invention. The present invention  
CC relates to the in vivo production of recombinant nucleic acid sequences.  
CC The method comprises expressing in a cell, two transcripts, one  
CC containing a first exon and first intron component, and a second  
CC transcript comprising a second intron component and a second exon.  
CC Transcript 1 and transcript 2 are allowed to trans-splice, forming a  
CC product containing exon 1 and exon 2, but not the intron components. The  
CC invention makes use of the ability of intronic sequences derived from  
CC group I or group II introns to mediate specific cleavage and ligation of  
CC discontinuous nucleic acid molecules. The method is used to produce  
CC recombinant nucleic acids, their products, or ribozymes in vivo. The  
CC method is preferably used for the preparation of combinatorial gene  
CC libraries in which the order and composition of exons are random. The  
CC method is also used for ordered gene assembly, or for the assembly of  
CC genes ordered at some exons but randomized at others. New genes can be  
CC selected rapidly and efficiently, and a very wide range of exons may be  
CC trapped.  
XX  
SQ Sequence 48 BP; 10 A; 10 C; 12 G; 16 T; 0 other;

Query Match 66.0%; Score 13.2; DB 21; Length 48;  
Best Local Similarity 83.3%; Pred. No. 1e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gtaagcagctgccattat 19  
| | | | | | | | | |  
Db 32 GAAAGCGGCGCCATTAT 15

Search completed: October 2, 2001, 16:18:39  
Job time: 15483 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:00:54 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-10

Perfect score: 20

Sequence: 1 ggtaagcagctgccattatt 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est33:\*  
111: gb\_est34:\*  
112: gb\_est35:\*  
113: gb\_est36:\*  
114: gb\_est37:\*  
115: gb\_est38:\*  
116: gb\_est39:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	15.2	76.0	50	107	AU103462	AU103462 AU103462
C 2	13.2	66.0	41	247	AZ634745	AZ634745 IM0490012
C 3	12.4	62.0	45	12	AL794942	AL794942 Tr05g03.r
C 4	12.4	61.0	28	258	TA355F030	TA355F030 T. brucei
C 5	12.2	61.0	30	258	TA272C090	TA272C090 T. brucei
C 6	12.2	61.0	40	12	AA789284	AA789284 aJ28c05.s
C 7	11.8	59.0	31	249	AZ798762	AZ798762 2M0055C18
C 8	11.8	59.0	33	241	AZ3111284	AZ3111284 IM0026G08
C 9	11.8	59.0	41	12	AA829551	AA829551 of09h01.s
C 10	11.8	59.0	45	188	TI7561	TI7561 mps v160 Th
C 11	11.8	59.0	50	10	AA662608	AA662608 nrl16a12.s
C 12	11.8	59.0	50	107	AU102666	AU102666 AU102666
C 13	11.8	59.0	50	107	AU103485	AU103485 AU103485
C 14	11.8	59.0	50	107	AU103486	AU103486 AU103486
C 15	11.8	59.0	50	107	AU103487	AU103487 AU103487
C 16	11.8	59.0	50	107	AU103488	AU103488 AU103488
C 17	11.8	59.0	50	107	AU103489	AU103489 AU103489
C 18	11.8	59.0	50	107	AU103493	AU103493 AU103493
C 19	11.6	58.0	40	106	AU006965	AU006965 AU006965
C 20	11.6	58.0	43	244	AZ476191	AZ476191 IM0294H17
C 21	11.6	58.0	44	249	AZ797074	AZ797074 2M0053P04
C 22	11.6	58.0	48	244	AZ492766	AZ492766 IM0327L03
C 23	11.6	58.0	49	13	AA914273	AA914273 YY99a03.r
C 24	11.6	58.0	50	107	AU105748	AU105748 AU105748
C 25	11.4	57.0	31	156	D18234	D18234 MUGSG00513
C 26	11.4	57.0	34	249	AZ785014	AZ785014 2M0028D03
C 27	11.4	57.0	47	143	BF036368	BF036368 601460534
C 28	11.4	57.0	50	107	AU103473	AU103473 AU103473
C 29	11.2	56.0	19	244	AZ481008	AZ481008 IM0302N15
C 30	11.2	56.0	20	249	AZ768476	AZ768476 IM0568D08
C 31	11.2	56.0	22	246	AZ598225	AZ598225 IM0412K22
C 32	11.2	56.0	33	147	BF346987	BF346987 602021958
C 33	11.2	56.0	39	250	AZ834606	AZ834606 2M0117G22
C 34	11.2	56.0	43	241	AZ329485	AZ329485 IM0053B17
C 35	11.2	56.0	50	20	AI476044	AI476044 t197b11.x
C 36	11.2	56.0	50	107	AU107071	AU107071 AU107071
C 37	11.2	56.0	50	107	AU107430	AU107430 AU107430
C 38	11.2	56.0	50	107	AU107431	AU107431 AU107431
C 39	11.2	56.0	50	173	BG077090	BG077090 H3010F02-
C 40	11	55.0	26	156	D21048	D21048 HUMGS02033
C 41	11	55.0	36	243	AZ425719	AZ425719 IM0205N21
C 42	11	55.0	36	244	AZ482058	AZ482058 IM0306J21
C 43	11	55.0	40	247	AZ644893	AZ644893 IM0510H05
C 44	11	55.0	41	258	TA386D04P	TA386D04P T. brucei
C 45	11	55.0	46	192	AK019062	AK019062 Mus muscu

## ALIGNMENTS

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	15.2	76.0	50	107	AU103462	AU103462 AU103462
C 2	13.2	66.0	41	247	AZ634745	AZ634745 IM0490012
C 3	12.4	62.0	45	12	AL794942	AL794942 Tr05g03.r
C 4	12.4	61.0	28	258	TA355F030	TA355F030 T. brucei
C 5	12.2	61.0	30	258	TA272C090	TA272C090 T. brucei
C 6	12.2	61.0	40	12	AA789284	AA789284 aJ28c05.s
C 7	11.8	59.0	31	249	AZ798762	AZ798762 2M0055C18
C 8	11.8	59.0	33	241	AZ3111284	AZ3111284 IM0026G08
C 9	11.8	59.0	41	12	AA829551	AA829551 of09h01.s
C 10	11.8	59.0	45	188	TI7561	TI7561 mps v160 Th
C 11	11.8	59.0	50	10	AA662608	AA662608 nrl16a12.s
C 12	11.8	59.0	50	107	AU102666	AU102666 AU102666
C 13	11.8	59.0	50	107	AU103485	AU103485 AU103485
C 14	11.8	59.0	50	107	AU103486	AU103486 AU103486
C 15	11.8	59.0	50	107	AU103487	AU103487 AU103487
C 16	11.8	59.0	50	107	AU103488	AU103488 AU103488
C 17	11.8	59.0	50	107	AU103489	AU103489 AU103489
C 18	11.8	59.0	50	107	AU103493	AU103493 AU103493
C 19	11.6	58.0	40	106	AU006965	AU006965 AU006965
C 20	11.6	58.0	43	244	AZ476191	AZ476191 IM0294H17
C 21	11.6	58.0	44	249	AZ797074	AZ797074 2M0053P04
C 22	11.6	58.0	48	244	AZ492766	AZ492766 IM0327L03
C 23	11.6	58.0	49	13	AA914273	AA914273 YY99a03.r
C 24	11.6	58.0	50	107	AU105748	AU105748 AU105748
C 25	11.4	57.0	31	156	D18234	D18234 MUGSG00513
C 26	11.4	57.0	34	249	AZ785014	AZ785014 2M0028D03
C 27	11.4	57.0	47	143	BF036368	BF036368 601460534
C 28	11.4	57.0	50	107	AU103473	AU103473 AU103473
C 29	11.2	56.0	19	244	AZ481008	AZ481008 IM0302N15
C 30	11.2	56.0	20	249	AZ768476	AZ768476 IM0568D08
C 31	11.2	56.0	22	246	AZ598225	AZ598225 IM0412K22
C 32	11.2	56.0	33	147	BF346987	BF346987 602021958
C 33	11.2	56.0	39	250	AZ834606	AZ834606 2M0117G22
C 34	11.2	56.0	43	241	AZ329485	AZ329485 IM0053B17
C 35	11.2	56.0	50	20	AI476044	AI476044 t197b11.x
C 36	11.2	56.0	50	107	AU107071	AU107071 AU107071
C 37	11.2	56.0	50	107	AU107430	AU107430 AU107430
C 38	11.2	56.0	50	107	AU107431	AU107431 AU107431
C 39	11.2	56.0	50	173	BG077090	BG077090 H3010F02-
C 40	11	55.0	26	156	D21048	D21048 HUMGS02033
C 41	11	55.0	36	243	AZ425719	AZ425719 IM0205N21
C 42	11	55.0	36	244	AZ482058	AZ482058 IM0306J21
C 43	11	55.0	40	247	AZ644893	AZ644893 IM0510H05
C 44	11	55.0	41	258	TA386D04P	TA386D04P T. brucei
C 45	11	55.0	46	192	AK019062	AK019062 Mus muscu

## JOURNAL COMMENT

mrnas using full-length enriched and 5'-end enriched cDNA libraries  
 Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: ysuzuki@ims.u-tokyo.ac.jp  
 Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano  
 S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

source  
 1..50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HEP14884"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
 BASE COUNT 9 a 15 c 9 g 17 t  
 ORIGIN

Query Match 76.0%; Score 15.2; DB 107; Length 50;  
 Best Local Similarity 85.0%; Pred. No. 1.7e+03;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtaagcagctgcattatt 20  
 ||| ||||| ||| |||  
 Db 41 GGTGAGCAGCTGCTATAATT 22

## RESULT 2

AZ634745/c 41 bp DNA GSS 13-DEC-2000  
 LOCUS  
 DEFINITION  
 1M0490012R Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0490012 R, DNA sequence.

ACCESSION  
 AZ634745  
 VERSION  
 AZ634745.1 GI:11756935  
 KEYWORDS  
 GSS.

## SOURCE

ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 41)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D. Weiss, R.

## REFERENCE

AUTHORS  
 TITLE  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0490 row: 0 column: 12  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 41.

## FEATURES

source  
 1..41  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0490012"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, P-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson



TA272C09Q/c  
 LOCUS TA272C09Q 30 bp DNA GSS 13-DEC-2000  
 DEFINITION T. brucei sheared genomic DNA clone 272c09, reverse sequence, genomic survey sequence.  
 ACCESSION AL487818  
 VERSION AL487818.1 GI:11852489  
 KEYWORDS GSS.  
 ORGANISM Trypanosoma brucei.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 REFERENCE 1 (bases 1 to 30)  
 AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk  
 COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU27/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).  
 Email: nelsayed@tigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/projects/T-brucei/>.  
 Location/Qualifiers

FEATURES  
 source  
 1. .30  
 /organism="Trypanosoma brucei"  
 /strain="TREU27"  
 /db\_xref="taxon:5691"  
 /clone="272c09"  
 9 a 10 c 3 g 8 t

BASE COUNT  
 ORIGIN  
 9 a 10 c 3 g 8 t

Query Match 61.0%; Score 12.2; DB 258; Length 30;  
 Best Local Similarity 82.4%; Pred. No. 5.2e+04;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 taagcagctgccattat 19  
 | ||||| ||||| |||  
 Db 24 TGAGCAGATGCCAGTAT 8

RESULT 6  
 AA789284/c  
 LOCUS AA789284 40 bp mRNA EST 31-DEC-1998  
 DEFINITION aj28c05.s1 Soares\_testis\_NHT Homo sapiens cDNA clone 1391624 3' similar to TR:P91373 P91373 SIMILARITY TO RAT TEST PROTEIN. ;, mRNA sequence.  
 ACCESSION AA789284  
 VERSION AA789284.1 GI:2849404  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 40)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cyapbs-f@mail.nih.gov  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: MCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Insert Length: 1216 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 1.

FEATURES  
 source  
 1. .40  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="1391624"  
 /clone\_lib="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 11 a 9 c 12 g 8 t

BASE COUNT  
 ORIGIN  
 11 a 9 c 12 g 8 t

Query Match 61.0%; Score 12.2; DB 12; Length 40;  
 Best Local Similarity 82.4%; Pred. No. 5.5e+04;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtaagcagctgccatt 17  
 ||||| ||||| |||||  
 Db 18 GGTACACAGCTGGCATT 2

RESULT 7  
 AZ798762  
 LOCUS AZ798762 31 bp DNA GSS 16-FEB-2001  
 DEFINITION 2M055C18R Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUC2M055C18 R, DNA sequence.

ACCESSION AZ798762  
 VERSION AZ798762.1 GI:12949191  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 31)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A., and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0055 row: C column: 18  
 Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends  
High quality sequence stop: 31.  
Location/Qualifiers  
1. .31

## FEATURES

source

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0055C18"  
/clone\_lib="Mouse 10kb plasmid UUGC1m library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

8 a 6 c 8 g 9 t  
BASE COUNT  
ORIGIN

Query Match 59.0%; Score 11.8; DB 249; Length 31;  
Best Local Similarity 86.7%; Pred. NO. 8.4e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gcagctgccattatt 20  
||||| || |||||

Db 6 GCAGCAGCAATTATT 20

RESULT 8  
A2311284/c

LOCUS A2311284 33 bp DNA GSS 29-SEP-2000  
DEFINITION 1M0026G08R Mouse 10kb plasmid UUGC1m library Mus musculus genomic clone UUGC1M0026G08 R, DNA sequence.

ACCESSION A2311284

VERSION A2311284.1 GI:10354094

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 33)

REFERENCE

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5605

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0026 row: G column: 08

Seq primer: CACACAGGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 33.

Location/Qualifiers

## FEATURES

source

1. .33  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0026G08"  
/clone\_lib="Mouse 10kb plasmid UUGC1m library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

6 a 10 c 9 g 8 t  
BASE COUNT  
ORIGIN

Query Match 59.0%; Score 11.8; DB 241; Length 33;  
Best Local Similarity 86.7%; Pred. NO. 8.5e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggttaagcagctgcc 15  
||||| ||||| |||

Db 29 GGTAAAGTAGCTCCCA 15

RESULT 9  
AA829551/c

LOCUS AA829551 41 bp mRNA EST 29-APR-1998

DEFINITION of09h01.s1 NCI-CGAP\_Col2 Homo sapiens cDNA clone IMAGE:142057 3' similar to gb:M33552 LYMPHOCYTE-SPECIFIC PROTEIN LSP1 (HUMAN); mRNA sequence.

ACCESSION AA829551

VERSION AA829551.1 GI:2902650

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 41)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Stratagene, Inc.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
 Insert Length: 660 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 1.

#### FEATURES

source  
 1. .41  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1420657"  
 /clone\_lib="NCI\_CGAP\_Col2"  
 /sex="mixed"  
 /tissue\_type="colon tumor"

/lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: colon; Vector: Bluescript SK-; Site\_1: EcorI  
 ; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
 dT. Pooled colon tumors. 5' adaptor sequence: 5'  
 GATTCGGACGAG 3' 3' adaptor sequence: 5'  
 CTCGATTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

#### BASE COUNT

10 a .6 c 11 g 14 t

#### Query Match

Best Local Similarity 59.0%; Score 11.8; DB 12; Length 41;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

#### QY

3 taagcagtcgacct 17

|||||

#### Db

19 TAGCAGCCACCAT 5

#### RESULT 10

##### TI17561/c

##### LOCUS

TI17561 45 bp mRNA EST 06-JUN-1994

mps v160 The blue guys library Saccharomyces cerevisiae cDNA  
 sequence upstream of LacZ fusion similar to ORFW, M91073, mRNA  
 sequence.

##### ACCESSION

TI17561.1 GI:458583

##### KEYWORDS

##### SOURCE

Saccharomyces cerevisiae.

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 45)

Burns, N., Grimwade, B., Ross-Macdonald, P.B., Choi, E.-Y., Finberg, K.,

Roeder, G.S. and Snyder, M.

Large-scale analysis of gene expression, protein localization and

gene disruption in Saccharomyces cerevisiae

Genes Dev. 8, 1087-1105 (1994)

95011603

Contact: Snyder M

Department of Biology

Yale University

New Haven CT 06520-8103

Tel: 2034326139

Fax: 2034326161

Email: snympcyalevm.ycc.yale.edu

LacZ fusion; Vegetative expression; Beta-gal fusion localization

pattern:

50 cytoplasmic spots; Disruption phenotype: inviable, unbudded

arrest; Fusion: codon12 of ORF next to BRFL. Sequence below near

or adjacent to lacZ.

Seq primer: LacZ sequences in transposon.

#### FEATURES

source  
 1. .45  
 /organism="Saccharomyces cerevisiae"  
 /db\_xref="taxon:4932"  
 /clone\_lib="The blue guys library"  
 /lab\_host="E.coli"

/note="Vector: pRECtmn; A yeast genomic DNA library was

prepared in the vector pHS56, and subjected to transposon

mutagenesis with mTn3. This mini-transposon carries lacZ

sequences that lack an initiation codon; expression of  
 lacZ is only provided by in frame fusion to yeast coding  
 sequence. The yeast genomic DNA carrying the transposon  
 was excised from pHS56 and transplanted back onto the yeast  
 chromosome. Yeast colonies expressing lacZ were screened  
 for in a color assay. A plasmid containing the genomic  
 DNA/lacZ fusion junction was recovered from each  
 individual yeast colony that expressed lacZ activity.  
 These recovered plasmids comprise 'the blue guys library'.  
 The fusion junction was then sequenced to identify the  
 expressed ORF upstream of the fusion."

BASE COUNT 14 a 13 c 8 g 10 t

#### ORIGIN

##### Query Match

Best Local Similarity 59.0%; Score 11.8; DB 188; Length 45;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

#### QY

6 gcagctgccattatt 20

|||||

#### Db

15 GCAGCTGCCACTAAT 1

#### RESULT 11

##### AA662608

##### LOCUS

AA662608 50 bp mRNA EST 12-NOV-1997

nr16a12.s1 NCI\_CGAP\_Ew1 Homo sapiens cDNA clone IMAGE:1168126  
 similar to SW:RL24\_HUMAN P38663 60S RIBOSOMAL PROTEIN L24. ;, mRNA  
 sequence.

##### ACCESSION

AA662608

##### VERSION

AA662608.1

##### KEYWORDS

##### SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 50)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,

Ph.D.

CDNA Library Preparation: David B. Krizman, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham.

#### FEATURES

source  
 1. .50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1168126"  
 /clone\_lib="NCI\_CGAP\_Ew1"  
 /tissue\_type="Ewing's sarcoma"  
 /lab\_host="DH10B"

/note="Vector: pAMP10; mRNA made from Ewing's sarcoma,

cDNA made by oligo-dT priming. Non-directionally cloned.

Size-selected on agarose gel, average insert size 600 bp.

Reference: Krizman et al. (1996) Cancer Research

56:5380-5383.

#### BASE COUNT

9 a 11 c 14 g 16 t

#### ORIGIN

##### Query Match

Best Local Similarity 59.0%; Score 11.8; DB 10; Length 50;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 taagcagctgccatt 17  
 | | | | | | | | | |  
 Db 11 TTAGCAGCAGCCATT 25

## RESULT 12

AU102666 50 bp mRNA EST 05-APR-2001  
 LOCUS AU102666 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 DEFINITION ADRA01276, mRNA sequence.  
 ACCESSION AU102666  
 VERSION AU102666.1 GI:13552187  
 EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 50)  
 AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,K., Suyama,A. and Sugano,S.  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo

## TITLE

Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

## JOURNAL

Unpublished (2001)

## COMMENT

Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

Source 1..50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="ADRA01276"  
 /clone\_lib="Sugano Homo sapiens cDNA library"

## BASE COUNT

11 a 17 c 12 g 10 t

## Query Match

Best Local Similarity 59.0%; Score 11.8; DB 107; Length 50;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 aagcagctgccatta 18  
 | | | | | | | | | |  
 Db 3 AAGCAGCTGGGATTA 17

## RESULT 13

AU103485 50 bp mRNA EST 05-APR-2001  
 LOCUS AU103485 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 DEFINITION COL04284, mRNA sequence.

## ACCESSION

AU103485

## VERSION

AU103485.1 GI:13553006

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 50)

## AUTHORS

Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,K., Suyama,A. and Sugano,S.  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo

## TITLE

Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

## JOURNAL

Unpublished (2001)

## COMMENT

Contact: Yutaka Suzuki  
 Department of Virology

Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: ysuzuki@ims.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

Location/Qualifiers  
 1..50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="COL04284"  
 /clone\_lib="Sugano Homo sapiens cDNA library"

## BASE COUNT

9 a 17 c 12 g 12 t

## ORIGIN

Query Match 59.0%; Score 11.8; DB 107; Length 50;

Best Local Similarity 86.7%; Pred. NO. 9.le+04;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggtaagcagctgccca 15  
 | | | | | | | | | |  
 Db 38 GGCAGGCAGCTGCCA 24

## RESULT 14

AU103486/c 50 bp mRNA EST 05-APR-2001  
 LOCUS AU103486 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 DEFINITION COLF4166, mRNA sequence.

## ACCESSION

AU103486

## VERSION

AU103486.1 GI:13553007

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 50)  
 AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.

## TITLE

Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

## JOURNAL

Unpublished (2001)

## COMMENT

Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

Location/Qualifiers  
 1..50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="COLF4166"  
 /clone\_lib="Sugano Homo sapiens cDNA library"

## BASE COUNT

11 a 19 c 11 g 9 t

## ORIGIN

Query Match 59.0%; Score 11.8; DB 107; Length 50;

Best Local Similarity 86.7%; Pred. NO. 9.le+04;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggtaagcagctgccca 15  
 | | | | | | | | | |  
 Db 38 GGCAGGCAGCTGCCA 24

## RESULT 15

AU103487/c

LOCUS AU103487 50 bp mRNA EST 05-APR-2001  
DEFINITION AU103487 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
COLF6739, mRNA sequence.  
ACCESSION AU103487  
VERSION AU103487.1 GI:13553008  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
,K., Suyama,A. and Sugano,S.  
TITLE Fine structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
JOURNAL Unpublished (2001)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: ysuzuki@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
,S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES  
source 1..50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="COLF6739"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 12 a 19 c 11 g 8 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 107; Length 50;  
Best Local Similarity 86.7%; Pred. No. 9.1e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ggtaagcagctgcc 15  
||| |  
Db 40 GGGAGGCAGCTGCC 26

Search completed: October 2, 2001, 15:00:56  
Job time: 10825 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:45 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100b-11  
Perfect score: 20  
Sequence: 1 agtaccagtgagtcttag 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_ba3:\*

4: gb\_in1:\*

5: gb\_in2:\*

6: gb\_in3:\*

7: gb\_cm:\*

8: gb\_ov:\*

9: gb\_pat1:\*

10: gb\_pat2:\*

11: gb\_ph:\*

12: gb\_pl1:\*

13: gb\_pl2:\*

14: gb\_pl3:\*

15: gb\_pl4:\*

16: em\_ba1:\*

17: em\_ba2:\*

18: em\_fun:\*

19: em\_htgo\_hum:\*

20: em\_htgo\_inv:\*

21: em\_htgo\_rsd:\*

22: em\_htg\_hum1:\*

23: em\_htg\_hum2:\*

24: em\_htg\_hum3:\*

25: em\_htg\_hum4:\*

26: em\_htg\_hum5:\*

27: em\_htg\_hum6:\*

28: em\_htg\_hum7:\*

29: em\_htg\_hum8:\*

30: em\_htg\_inv1:\*

31: em\_htg\_inv2:\*

32: em\_htg\_other:\*

33: em\_htg\_rsd:\*

34: em\_hum1:\*

35: em\_hum2:\*

36: em\_hum3:\*

37: em\_hum4:\*

38: em\_hum5:\*

39: em\_hum6:\*

40: em\_hum7:\*

41: em\_in:\*

42: em\_or:\*

43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_rsd:\*

50: em\_sv:\*

51: em\_un:\*

52: em\_vl:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sv:\*

57: gb\_un:\*

58: gb\_vl:\*

59: gb\_vl2:\*

60: gb\_htg1:\*

61: gb\_htg2:\*

62: gb\_htg3:\*

63: gb\_htg4:\*

64: gb\_htg5:\*

65: gb\_htg6:\*

66: gb\_htg7:\*

67: gb\_htg8:\*

68: gb\_htg9:\*

69: gb\_htg10:\*

70: gb\_htg11:\*

71: gb\_htg12:\*

72: gb\_htg13:\*

73: gb\_htg14:\*

74: gb\_htg15:\*

75: gb\_htg16:\*

76: gb\_htg17:\*

77: gb\_htg18:\*

78: gb\_htg19:\*

79: gb\_htg20:\*

80: gb\_htg21:\*

81: gb\_htg22:\*

82: gb\_htg23:\*

83: gb\_htg24:\*

84: gb\_htg25:\*

85: gb\_pr1:\*

86: gb\_pr2:\*

87: gb\_pr3:\*

88: gb\_pr4:\*

89: gb\_pr5:\*

90: gb\_pr6:\*

91: gb\_pr7:\*

92: gb\_pr8:\*

93: gb\_pr9:\*

94: gb\_rsd:\*

95: gb\_rsd2:\*

96: gb\_in4:\*

97: gb\_pr10:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	13.4	67.0	21	10	I30544 Sequence 7
C 2	12.8	64.0	29	10	I75320 Sequence 69
C 3	12.8	64.0	34	9	AR091705 Sequence
C 4	12.8	64.0	48	9	AR075823 Sequence
C 5	12.8	64.0	48	10	E30456 Sequence
C 6	12.4	62.0	18	9	A34802 Method for
C 7	12.4	62.0	22	9	AR067320 Sequence
C 8	12.2	61.0	44	9	AR035220 Sequence

9 12.2 61.0 44 9 AR035241 Sequence  
 10 12 60.0 20 9 AR060552 Sequence  
 11 12 60.0 39 9 A28774 Oligonucleo  
 12 11.8 59.0 41 9 A28774 Sequence  
 13 11.6 58.0 23 9 A49393 Sequence 11  
 14 11.6 58.0 26 9 A97534 Sequence 34  
 15 11.6 58.0 26 9 AR013888 Sequence  
 16 11.6 58.0 26 9 AR033842 Sequence  
 17 11.6 58.0 26 9 AR042502 Sequence  
 18 11.6 58.0 26 9 AR058382 Sequence  
 19 11.6 58.0 26 9 AR088208 Sequence  
 20 11.6 58.0 26 9 AR089875 Sequence  
 21 11.6 58.0 35 9 A17379 Nucleotide  
 22 11.6 58.0 40 9 AR013900 Sequence  
 23 11.6 58.0 40 9 AR033854 Sequence  
 24 11.6 58.0 40 9 AR042514 Sequence  
 25 11.6 58.0 40 9 AR058394 Sequence  
 26 11.6 58.0 40 9 AR088220 Sequence  
 27 11.6 58.0 42 9 AR013872 Sequence  
 28 11.6 58.0 42 9 AR021426 Sequence  
 29 11.6 58.0 42 9 AR033826 Sequence  
 30 11.6 58.0 42 9 AR042486 Sequence  
 31 11.6 58.0 42 9 AR042988 Sequence  
 32 11.6 58.0 42 9 AR058366 Sequence  
 33 11.6 58.0 42 9 AR088192 Sequence  
 34 11.6 58.0 42 10 I43967 Sequence 58  
 35 11.6 58.0 42 10 I62979 Sequence 10  
 36 11.6 58.0 42 10 I88732 Sequence 10  
 37 11.4 57.0 20 10 E28760 Antitumor d  
 38 11.4 57.0 24 9 A25691  
 39 11.4 57.0 24 9 AR051366  
 40 11.4 57.0 30 10 I36150 Sequence 34  
 41 11.4 57.0 38 9 AR061462 Sequence  
 42 11.4 57.0 38 9 AR108361 Sequence  
 43 11.4 57.0 38 10 I16318 Sequence 14  
 44 11.4 57.0 38 10 I66804 Sequence 14  
 45 11.4 57.0 38 10 I84898 Sequence 14

## ALIGNMENTS

RESULT 1  
 130544/c  
 LOCUS Sequence 21 bp DNA  
 DEFINITION Sequence 7 from patent US 5580969.  
 ACCESSION 130544  
 VERSION 130544.1 GI:1821335  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 21)  
 AUTHORS Hoke,G.D., Bradley,M.O., Williams,T.J. and Lee,C.  
 TITLE Antisense oligonucleotides directed against human ICAM-1 RNA  
 JOURNAL Patent: US 5580969-A 7 03-DEC-1996;  
 FEATURES Location/Qualifiers  
 1..21  
 /organism="unknown"  
 BASE COUNT 5 a 9 c 5 g 2 t  
 ORIGIN  
 Query Match 67.0%; Score 13.4; DB 10; Length 21;  
 Best Local Similarity 93.3%; Pred. No. 6.4e+03;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 gtaccacggtagtc 16  
 ||| ||||| |||||  
 Db 19 GTTCCAGGTGAGTC 5  
 RESULT 2  
 130544/c  
 LOCUS Sequence 21 bp DNA  
 DEFINITION Sequence 7 from patent US 5580969.  
 ACCESSION 130544  
 VERSION 130544.1 GI:1821335  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 21)  
 AUTHORS Hoke,G.D., Bradley,M.O., Williams,T.J. and Lee,C.  
 TITLE Antisense oligonucleotides directed against human ICAM-1 RNA  
 JOURNAL Patent: US 5580969-A 7 03-DEC-1996;  
 FEATURES Location/Qualifiers  
 1..21  
 /organism="unknown"  
 BASE COUNT 5 a 9 c 5 g 2 t  
 ORIGIN  
 Query Match 67.0%; Score 13.4; DB 10; Length 21;  
 Best Local Similarity 93.3%; Pred. No. 6.4e+03;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 gtaccacggtagtc 16  
 ||| ||||| |||||  
 Db 19 GTTCCAGGTGAGTC 5

I75320/c  
 LOCUS Sequence 29 bp DNA  
 DEFINITION Sequence 69 from patent US 5689052.  
 ACCESSION I75320  
 VERSION I75320.1 GI:3011461  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 29)  
 AUTHORS Brown,S.Marie, Dean,D.Allen, Fromm,M.Ernest and Sanders,P.Rigden.  
 TITLE Synthetic DNA sequences having enhanced expression in  
 monocotyledonous plants and method for preparation thereof  
 JOURNAL Patent: US 5689052-A 69 18-NOV-1997;  
 FEATURES Location/Qualifiers  
 1..29  
 /organism="unknown"  
 BASE COUNT 6 a 9 c 6 g 8 t  
 ORIGIN  
 Query Match 64.0%; Score 12.8; DB 10; Length 29;  
 Best Local Similarity 87.5%; Pred. No. 1.4e+04;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 agtaccacggtagtc 16  
 ||| ||||| |||||  
 Db 24 AGTGGCCAGGAGAGTC 9  
 RESULT 3  
 AR091705  
 LOCUS AR091705 34 bp DNA  
 DEFINITION Sequence 15 from patent US 5994505.  
 ACCESSION AR091705  
 VERSION AR091705.1 GI:10018459  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 34)  
 AUTHORS Ting,J.Pan-Yung and Chin,K.  
 TITLE Forms of class II MHC transactivator (CIITA)  
 JOURNAL Patent: US 5994505-A 15 30-NOV-1999;  
 FEATURES Location/Qualifiers  
 1..34  
 /organism="unknown"  
 BASE COUNT 7 a 11 c 10 g 6 t  
 ORIGIN  
 Query Match 64.0%; Score 12.8; DB 9; Length 34;  
 Best Local Similarity 87.5%; Pred. No. 1.4e+04;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 3 taccacggtagtc 18  
 ||||| |||||  
 Db 16 TACCAGCTGTCTCT 31

RESULT 4  
 AR075823/c  
 LOCUS AR075823 48 bp DNA  
 DEFINITION Sequence 6 from patent US 5958700.  
 ACCESSION AR075823  
 VERSION AR075823.1 GI:10002569  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 48)  
 AUTHORS Nadeau,J.G., Pitner,J.Bruce, Linn,C.Preston and Schram,J.L.  
 TITLE Detection of nucleic acids by fluorescence quenching

Renard,A. and Thiry,M. fish EUROGENTEC S.A location/Qualifiers 1. .18 /organism="synthetic construct" /db_xref="taxon:32630"		5 a	4 c	7 g	2 t	
Query Match Best Local Similarity 62.0%; Score 12.4; DB 9; Length 18; Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;						
Qy	5 cccagggtgagtcctt 18 					
Db	14 CCCAGGTGAGTCTT 1					
RESULT 7						
AR067320/c						
LOCUS	AR067320 22 bp DNA PAT 29-SEP-1999					
DEFINITION	Sequence 668 from patent US 5851760.					
ACCESSION	AR067320					
VERSION	AR067320.1 GI:5998542					
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 22)					
AUTHORS	Evans,G.A. and Smith,M.W.					
TITLE	Method for generation of sequence sampled maps of complex genomes					
JOURNAL	Patent: US 5851760-A 668 22-DEC-1998;					
FEATURES	Location/Qualifiers 1. .22 /organism="unknown"					
BASE COUNT	8 a 5 c 4 g 5 t					
ORIGIN						
Query Match 62.0%; Score 12.4; DB 9; Length 22; Best Local Similarity 92.9%; Pred. No. 2.6e+04; Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;						
Qy	4 acccaggtgagtcct 17 					
Db	14 ATCCAGGTGAGTCT 1					
RESULT 8						
AR035220/c						
LOCUS	AR035220 44 bp DNA PAT 29-SEP-1999					
DEFINITION	Sequence 36 from patent US 5871732.					
ACCESSION	AR035220					
VERSION	AR035220.1 GI:5951888					
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 44)					
AUTHORS	Burkly,L.C., Chisholm,P.L., Thomas,D.W., Rosa,M.D. and Rosa,J.J.					
TITLE	Anti-CD4 antibody homologs useful in prophylaxis and treatment of AIDS, ARC and HIV infection					
JOURNAL	Patent: US 5871732-A 36 16-FEB-1999;					
FEATURES	Location/Qualifiers 1. .44 /organism="unknown"					
BASE COUNT	11 a 12 c 14 g 7 t					
ORIGIN						

Query Match 61.0%; Score 12.2; DB 9; Length 44;  
Best Local Similarity 82.4%; Pred. No. 3.2e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 taccacggtgagtcctta 19  
| | | | | | | | | | |  
Db 27 TCCTCAGGTGAGTCCTA 11

## RESULT 9

AR035241 AR035241 44 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 59 from patent US 5871732.

ACCESSION AR035241  
VERSION AR035241.1 GI:5951909

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 44)

AUTHORS Burkly,L.C., Chisholm,P.L., Thomas,D.W., Rosa,M.D. and Rosa,J.J.

TITLE Anti-CD4 antibody homologs useful in prophylaxis and treatment of AIDS, ARC and HIV infection

JOURNAL Patent: US 5871732-A 59 16-FEB-1999;  
FEATURES Location/Qualifiers

1..44  
/organism="unknown"

BASE COUNT 7 a 14 c 12 g 11 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 44;  
Best Local Similarity 82.4%; Pred. No. 3.2e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 taccacggtgagtcctta 19  
| | | | | | | | | | |  
Db 22 TCCTCAGGTGAGTCCTA 38

## RESULT 10

AR060552 AR060552 20 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 53 from patent US 5840693.

ACCESSION AR060552  
VERSION AR060552.1 GI:5987002

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 20)

AUTHORS Eriksson,U., Olofsson,B., Alitalo,K. and Pajusola,K.

TITLE Vascular endothelial growth factor-B

JOURNAL Patent: US 5840693-A 53 24-NOV-1998;  
FEATURES Location/Qualifiers

1..20  
/organism="unknown"

BASE COUNT 5 a 4 c 6 g 5 t  
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.5e+04;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 caggtgagtcctt 18  
| | | | | | | | | | |  
Db 8 CAGGTGAGTCCTT 19

## RESULT 11

A28774

LOCUS A28774 39 bp DNA PAT 12-JUN-1995  
DEFINITION Oligonucleotide xol123.  
ACCESSION A28774  
VERSION A28774.1 GI:1248748  
KEYWORDS synthetic construct.  
SOURCE synthetic construct  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 39)  
AUTHORS

TITLE PROTEINS PREVENTING INTERACTION BETWEEN A FRAGMENT FC OF AN

IMMUNOGLOBULIN AND ITS RECEIVER, AND THERAPEUTIC USE, PARTICULARLY

IN TREATING AFFECTIONS LINKED TO THE HIV VIRUS

JOURNAL Patent: WO 9108301-A 6 13-JUN-1991;  
FEATURES Location/Qualifiers

1..39  
/organism="synthetic construct"

BASE COUNT 12 a 6 c 12 g 9 t  
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 39;  
Best Local Similarity 100.0%; Pred. No. 4.3e+04;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agtaccacggtg 12  
| | | | | | | | | | |  
Db 15 AGTACCCAGGTG 26

## RESULT 12

AR075947 AR075947 41 bp DNA PAT 30-AUG-2000  
DEFINITION Sequence 8 from patent US 5958713.

ACCESSION AR075947  
VERSION AR075947.1 GI:10002693

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 41)

AUTHORS Thastrup,O., Tullin,S.slashed.ren, Poulsen,L.Kongsbak and Bj.o  
slashed.rn,S.petersen.

TITLE Method of detecting biologically active substances by using green

JOURNAL fluorescent protein

FEATURES Patent: US 5958713-A 8 28-SEP-1999;  
Location/Qualifiers

1..41  
/organism="unknown"

BASE COUNT 8 a 9 c 12 g 12 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 41;  
Best Local Similarity 86.7%; Pred. No. 5.7e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 6 ccaggtgagtccttag 20  
| | | | | | | | | | |  
Db 20 CCGGTGAGTCATAG 34

## RESULT 13

A49393/c A49393 23 bp DNA PAT 07-MAR-1997  
DEFINITION Sequence 11 from Patent WO9607740.

ACCESSION A49393  
VERSION A49393.1 GI:2302870

KEYWORDS

SOURCE unidentified.

ORGANISM unidentified.

unclassified.

```

/organism="unknown"
BASE COUNT      7 a      6 c      9 g      4 t
ORIGIN

Query Match          58.0%; Score 11.6; DB 9; Length 26;
Best Local Similarity 77.8%; Pred. No. 7.7e+04;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3  taccacaggtgagtcttag 20
          | | | | | | | | | |
Dd      25  TCCTCAGGTGAGTCCTTG 8

Search completed: October 2, 2001, 15:56:45
Job time: 14169 sec

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:00:56 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-11

Perfect score: 20

Sequence: 1 agtaccagggtgagctcttag 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est25.\*
- 26: gb\_est26.\*
- 27: gb\_est27.\*
- 28: gb\_est28.\*
- 29: gb\_est29.\*
- 30: gb\_est30.\*
- 31: gb\_est31.\*
- 32: gb\_est32.\*
- 33: gb\_est33.\*
- 34: gb\_est34.\*
- 35: gb\_est35.\*
- 36: gb\_est36.\*
- 37: gb\_est37.\*
- 38: gb\_est38.\*
- 39: gb\_est39.\*
- 40: gb\_est40.\*
- 41: gb\_est41.\*
- 42: gb\_est42.\*
- 43: gb\_est43.\*
- 44: gb\_est44.\*
- 45: gb\_est45.\*
- 46: gb\_est46.\*
- 47: gb\_est47.\*

- 44: em\_esthum10.\*
- 45: em\_esthum11.\*
- 46: em\_esthum12.\*
- 47: em\_esthum13.\*
- 48: em\_esthum14.\*
- 49: em\_esthum15.\*
- 50: em\_esthum16.\*
- 51: em\_esthum17.\*
- 52: em\_esthum18.\*
- 53: em\_esthum19.\*
- 54: em\_esthum20.\*
- 55: em\_esthum21.\*
- 56: em\_esthum22.\*
- 57: em\_esthum23.\*
- 58: em\_esthum24.\*
- 59: em\_esthum25.\*
- 60: em\_esthum26.\*
- 61: em\_esthum27.\*
- 62: em\_esthum28.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estin5.\*
- 68: em\_estom1.\*
- 69: em\_estom2.\*
- 70: em\_estov1.\*
- 71: em\_estov2.\*
- 72: em\_estpl1.\*
- 73: em\_estpl2.\*
- 74: em\_estpl3.\*
- 75: em\_estpl4.\*
- 76: em\_estpl5.\*
- 77: em\_estpl6.\*
- 78: em\_estpl7.\*
- 79: em\_estpl8.\*
- 80: em\_estpl9.\*
- 81: em\_estpl10.\*
- 82: em\_estro1.\*
- 83: em\_estro2.\*
- 84: em\_estro3.\*
- 85: em\_estro4.\*
- 86: em\_estro5.\*
- 87: em\_estro6.\*
- 88: em\_estro7.\*
- 89: em\_estro8.\*
- 90: em\_estro9.\*
- 91: em\_estro10.\*
- 92: em\_estro11.\*
- 93: em\_estro12.\*
- 94: em\_estro13.\*
- 95: em\_estro14.\*
- 96: em\_estro15.\*
- 97: em\_estro16.\*
- 98: em\_estro17.\*
- 99: em\_estro18.\*
- 100: em\_estro19.\*
- 101: em\_estro20.\*
- 102: gb\_est25.\*
- 103: gb\_est26.\*
- 104: gb\_est27.\*
- 105: gb\_est28.\*
- 106: gb\_est29.\*
- 107: gb\_est30.\*
- 108: gb\_est31.\*
- 109: gb\_est32.\*
- 110: gb\_est33.\*
- 111: gb\_est34.\*
- 112: gb\_est35.\*
- 113: gb\_est36.\*
- 114: gb\_est37.\*
- 115: gb\_est38.\*
- 116: gb\_est39.\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: em_esthum29:*
170: em_esthum30:*
171: em_esthum31:*
172: em_esthum32:*
173: em_esthum33:*
174: em_estom3:*
175: em_estpill:*
176: em_estro21:*
177: em_estro22:*
178: em_estro23:*
179: em_hic:*
180: gb_est107:*
181: gb_est108:*
182: gb_est109:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0347 row: C column: 11  
 Seq primer: CGTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 25.  
 Location/Qualifiers  
 1. .25  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0347C11"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /notes="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydronamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gil4732114/gb/AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."  
 3 a 6 c 9 g 7 t  
 BASE COUNT  
 ORIGIN  
 Query Match 64.0%; Score 12.8; DB 245; Length 25;  
 Best Local Similarity 87.5%; Pred. No. 1.8e+04;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 ccacggtgagctcttag 20  
 Db 5 CTCAGGTCAGTCCTTG 20  
 RESULT 2  
 AZ666528  
 LOCUS 37 bp DNA GSS 14-DEC-2000  
 DEFINITION IM0548A18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0548A18 R, DNA sequence.  
 ACCESSION AZ666528  
 VERSION AZ666528  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 37)  
 REFERENCE  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 M., Rose,M., Stokes,R., Tingey,A., von Niederhausern,A.,  
 and Wright,D.,Weiss,R.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12.8	64.0	25	245	AZ506197 1M0347C11
2	12.6	63.0	37	247	AZ666528 1M0548A18
c 3	12.4	62.0	32	241	AZ303920 1M0003F19
4	12	60.0	27	159	L32037 HUMXP1F11B
5	11.8	59.0	50	107	AU105613 AU105613
c 6	11.6	58.0	46	250	AZ826374 AU105613
c 7	11.6	58.0	47	156	D38689 HUMC1264.Hu
8	11.2	56.0	31	249	AZ774468 2M0003017
c 9	11.2	56.0	33	249	AZ799299 2M0056M03
c 10	11.2	56.0	34	249	AZ763845 1M0559C04
c 11	11.2	56.0	43	244	AZ480568 1M0302M10
c 12	11.2	56.0	43	249	AZ806794 2M0069006
c 13	11.2	56.0	48	258	TA117F12P
c 14	11.2	56.0	50	107	AU106341
c 15	11	55.0	34	188	R84653
c 16	11	55.0	48	18	AI318658
c 17	11	55.0	50	107	AU102283
c 18	11	55.0	50	107	AU102992
c 19	11	55.0	50	107	AU105601
c 20	10.8	54.0	27	244	AZ495501 1M0331P21
c 21	10.8	54.0	28	241	AZ313467 1M0029P06
c 22	10.8	54.0	33	242	AZ382781 1M0140N08
c 23	10.8	54.0	38	245	AZ508030 1M0349P22
c 24	10.8	54.0	40	258	TA69H03Q
c 25	10.8	54.0	42	243	AZ427649
c 26	10.8	54.0	43	249	AZ785625
c 27	10.8	54.0	45	21	AI538057
c 28	10.8	53.0	50	107	AU105612
c 29	10.6	53.0	31	249	AZ770290
c 30	10.6	53.0	37	102	AI862676
c 31	10.6	53.0	37	247	AZ644324
c 32	10.6	53.0	42	11	AA741237
c 33	10.6	53.0	43	103	AI917489
c 34	10.6	53.0	43	187	S0655
c 35	10.6	53.0	46	159	H98043
c 36	10.6	53.0	46	249	AZ764208
c 37	10.6	53.0	48	250	AZ819513
c 38	10.6	53.0	49	168	BF740213
c 39	10.6	53.0	49	189	WI2708
c 40	10.6	53.0	50	107	AU106954
c 41	10.4	52.0	23	243	AZ433756
c 42	10.4	52.0	24	244	AZ478673
c 43	10.4	52.0	25	191	W95789
c 44	10.4	52.0	26	250	AZ827167
c 45	10.4	52.0	26	258	TA30H08Q

ALIGNMENTS

RESULT 1  
 AZ506197  
 LOCUS 25 bp DNA GSS 05-OCT-2000  
 DEFINITION 1M0347C11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0347C11 F, DNA sequence.  
 ACCESSION AZ506197  
 VERSION AZ506197.1 GI:10687513  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 25)  
 REFERENCE  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 M., Rose,M., Stokes,R., Tingey,A., von Niederhausern,A.,  
 and Wright,D.,Weiss,R.

**TITLE**  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
**JOURNAL**  
Unpublished (2000)  
**COMMENT**  
Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0548 row: A column: 18  
Seq primer: CACACAGGAACACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 37.  
Location/Qualifiers

**FEATURES**  
source

1. 37  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUCG1M003F19"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

**BASE COUNT** 14 a 7 c 8 g 8 t  
**ORIGIN**  
Query Match 63.0%; Score 12.6; DB 247; Length 37;  
Best Local Similarity 78.9%; Pred. No. 2.4e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 agtaccagggtgagctctta 19  
||| ||||| ||||| |||||  
Db 13 AGAACCCATGTGTAATGTTA 31

**RESULT** 3  
**LOCUS** A2303920 32 bp DNA GSS 29-SEP-2000  
**DEFINITION** IM0003F19R Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
clone UUCG1M0003F19 R, DNA sequence.  
**ACCESSION** A2303920  
**VERSION** A2303920.1 GI:10339372  
**KEYWORDS** GSS.  
**SOURCE** house mouse.  
**ORGANISM** Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE** 1 (bases 1 to 32)  
**AUTHORS** Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
**JOURNAL**  
Unpublished (2000)  
**COMMENT**  
Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0003 row: F column: 19  
Seq primer: CACACAGGAACACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 32.  
Location/Qualifiers

**FEATURES**  
source

1. 32  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUCG1M0003F19"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

**BASE COUNT** 5 a 10 c 9 g 8 t  
**ORIGIN**

Query Match 62.0%; Score 12.4; DB 241; Length 32;  
Best Local Similarity 92.9%; Pred. No. 3e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agtaccagggtgag 14  
||| ||||| |||||  
Db 31 AGGACCCAGGTGAG 18

**RESULT** 4  
**LOCUS** L32037 27 bp mRNA EST 02-AUG-1995  
**DEFINITION** HUMXP1F11B Human placenta Homo sapiens cDNA clone XP6G6B, mRNA  
sequence.  
**ACCESSION** L32037  
**VERSION** L32037.1 GI:927081  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 27)  
**AUTHORS** Lee, C.-C., Yazdani, A., Wehnert, M., Bailey, J., Couch, L., Xiong, M.,

Coolbaugh, M.I., Chinault, C.A., Baldini, A., Lindsay, E.A., Zhao, Z.-Y.  
and Caskey, C.T.H.  
Isolation of chromosome-specific genes by reciprocal probing of  
arrayed cDNAs and cosmid libraries  
Hum. Mol. Genet. 4, 1373-1380 (1995)  
96090257  
Contact: Caskey, C.T.H.

## TITLE

Isolation of chromosome-specific genes by reciprocal probing of  
arrayed cDNAs and cosmid libraries

## JOURNAL

Hum. Mol. Genet. 4, 1373-1380 (1995)

## MEDLINE

96090257

## COMMENT

Contact: Caskey, C.T.H.

## FEATURES

source  
1. .27  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="xq27.3-q28"  
/clone="XP6G6B"  
/notes="Arrayed cDNAs and cosmid libraries from human  
placental tissue"

7 a 4 c 7 g 9 t

## BASE COUNT

## ORIGIN

Query Match 60.0%; Score 12; DB 159; Length 27;  
Best Local Similarity 75.0%; Pred. No. 4.8e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 agtaccagggtgagctttag 20

||||| ||| ||| |||

Db 5 AATACCTGATGCGCTTTAG 24

## RESULT 5

## AUI05613

## LOCUS

AUI05613 50 bp mRNA EST 05-APR-2001  
DEFINITION HRP22150, mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 50)  
Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata  
H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo  
K., Suyama, A. and Sugano, S.  
Fine Structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
Unpublished (2001)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano  
S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

1. .50  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP22150"  
/clone\_lib="Sugano Homo sapiens cDNA library"

6 a 7 c 22 g 15 t

## BASE COUNT

## ORIGIN

Query Match 59.0%; Score 11.8; DB 107; Length 50;  
Best Local Similarity 86.7%; Pred. No. 6.5e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 6 ccagggtgagctttag 20

||||| ||| ||| |||

Db 13 CCGGGTGAGTCTTAG 27

## RESULT 6

## A2826374/c

## LOCUS

A2826374 46 bp DNA GSS 20-FEB-2001  
DEFINITION 2M0102K06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M0102K06 F, DNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 46)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0102 row: K column: 06  
Seq primer: CGTGTAAACGACGGCCAGT  
Class: plasmid ends  
High quality sequence stop: 46.

Location/Qualifiers  
1. .46  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0102K06"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/notes="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi14732114.gb|AF129072.1), a copy-number  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

17 a 14 c 9 g 6 t

## BASE COUNT

## ORIGIN

Query Match 58.0%; Score 11.6; DB 250; Length 46;  
Best Local Similarity 77.8%; Pred. No. 8.2e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 agtaccagggtgagctt 18

||||| ||| ||| |||

```

Db 21 AGTATCAGGTGCTGCTT 4

RESULT 7
LOCUS D38689 47 bp mRNA EST 31-MAY-1995
DEFINITION HUMC1264 Human chromosome 8 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION D38689
VERSION D38689.1 GI:807747
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 47)
AUTHORS Kovama,K., Sufo,K. and Nakamura,Y.
TITLE Isolation of 115 human chromosome 8-specific expressed-sequence
tags by exon amplification
JOURNAL Genomics 26, 245-253 (1995)
MEDLINE 95324915
COMMENT Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433
Email: yusuke@ims.u-tokyo.ac.jp.
Location/Qualifiers
1. 47
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
/clone.lib="Human chromosome 8"
/note="Exon-like sequence on chromosome 8; exon-trapping
method"

BASE COUNT 21 a 9 c 9 g 8 t
ORIGIN

Query Match 58.0%; Score 11.6; DB 156; Length 47;
Best Local Similarity 77.8%; Pred. No. 8.2e+04;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 taccagggtgagctcttag 20
||||| | | | |
Db 21 TACCCAGGTAATCTTCG 4

RESULT 8
LOCUS AZ774468 31 bp DNA GSS 16-FEB-2001
DEFINITION 2M0003017R Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUCG2M0003017 R, DNA sequence.
ACCESSION AZ774468
VERSION AZ774468.1 GI:12899949
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 31)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0003 row: O column: 17
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 31.
Location/Qualifiers
1. 31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG2M0003017"
/clone.lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
a 6 c 10 t

BASE COUNT 9 a 6 c 10 t
ORIGIN

Query Match 56.0%; Score 11.2; DB 249; Length 31;
Best Local Similarity 81.2%; Pred. No. 1.3e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 acccaggtgagctcta 19
||||| | | | |
Db 10 ACCCAGCTGGGTTTGA 25

RESULT 9
LOCUS AZ799299 33 bp DNA GSS 16-FEB-2001
DEFINITION 2M0056M03R Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUCG2M0056M03 R, DNA sequence.
ACCESSION AZ799299
VERSION AZ799299.1 GI:12950275
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 33)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA

```

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0056 row: M column: 03

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 33.

Location/Qualifiers

1. .33

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0056M03"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/notes="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gil14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

11 a 9 c 6 g 7 t

BASE COUNT  
ORIGIN

Query Match 56.0%; Score 11.2; DB 249; Length 33;

Best Local Similarity 81.2%; Pred. No. 1.3e+05;

Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agtaccacggtgagtc 16

||| |||| |||||

Db 33 AGTACCACGTTGAGTC 18

RESULT 10  
AZ763845/c

LOCUS AZ763845 34 bp DNA GSS 16-FEB-2001

DEFINITION IM0559C04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0559C04 R, DNA sequence.

ACCESSION

AZ763845

VERSION AZ763845.1 GI:12875290

KEYWORDS

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 34)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D.,Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0559 row: C column: 04

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 34.

Location/Qualifiers

1. .34

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0559C04"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/notes="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gil14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

9 a 12 c 9 g 4 t

BASE COUNT  
ORIGIN

Query Match 56.0%; Score 11.2; DB 249; Length 34;

Best Local Similarity 81.2%; Pred. No. 1.3e+05;

Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agtaccacggtgagtc 16

||| ||||| |||||

Db 28 AGTCCACGAGTGGGCC 13

RESULT 11  
AZ480568

LOCUS AZ480568 43 bp DNA GSS 04-OCT-2000

DEFINITION IM0302M10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0302M10 F, DNA sequence.

ACCESSION

AZ480568

VERSION AZ480568.1 GI:10641633

KEYWORDS

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 43)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D.,Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0302 row: M column: 10  
Seq primer: CGTTGTAACAGCAGCCAGT  
Class: plasmid ends  
High quality sequence stop: 43.

# FEATURES

Location/Qualifiers  
1. .43  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUC1M0302M10"  
/clone\_lib="Mouse 10kb plasmid UUC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 11 a 7 c 17 g 8 t  
ORIGIN  
Query Match 56.0%; Score 11.2; DB 244; Length 43;  
Best Local Similarity 81.2%; Pred. NO. 1.3e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 gtaccaggtgagctct 17  
||| |||| |||||  
Db 18 GTAGACAGGAGAGTCT 33

RESULT 12  
LOCUS A2806794 43 bp DNA GSS 20-FEB-2001  
DEFINITION 2M0069006F Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUC2M0069006 F, DNA sequence.  
ACCESSION A2806794  
VERSION A2806794.1 GI:12970500  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 43)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss

University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0069 row: O column: 06  
Seq primer: CGTTGTAACAGCAGCCAGT  
Class: plasmid ends  
High quality sequence stop: 43.

# FEATURES

Location/Qualifiers  
1. .43  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUC2M0069006"  
/clone\_lib="Mouse 10kb plasmid UUC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 8 a 5 c 9 g 21 t  
ORIGIN  
Query Match 56.0%; Score 11.2; DB 249; Length 43;  
Best Local Similarity 81.2%; Pred. NO. 1.3e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4 acccaggtgagctctta 19  
||| ||| |||||  
Db 34 ACCAAGGCAAGTCTTA 19

RESULT 13  
LOCUS TAI17F12P/c 48 bp DNA GSS 13-DEC-2000  
DEFINITION T. brucei sheared genomic DNA clone 117f12, forward sequence, genomic survey sequence.  
ACCESSION AL462744  
VERSION AL462744.1 GI:11832461  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
1 (bases 1 to 48)  
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.  
Direct Submission  
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

nhlsanger.ac.uk  
 Constructed at the Institute for Genomic Research (TIGR),  
 Rockville, MD. Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
 to give a tight size distribution (4 kb). The v + i method used for the library construction is  
 described in detail in Smith, H. and Venter, J.C. (Making small  
 insert libraries for whole genome shotgun sequencing projects. In  
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Barrell, Oxford University Press, 1999).  
 Email: nelsayed@tigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available  
 at <http://www.sanger.ac.uk/Projects/T-brucei/>.

## FEATURES

source

1. .48  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="117f12"

BASE COUNT 12 a 10 c 15 g 11 t

ORIGIN

Query Match 56.0%; Score 11.2; DB 258; Length 48;  
 Best Local Similarity 81.2%; Pred. No. 1.4e+05;  
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gtaccagggtgagctt 17

Db 16 GTACCCACGTCGCT 1  
 ||||| || ||||

RESULT 14

AU106341

LOCUS

DEFINITION AU106341 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone 05-APR-2001

ACCESSION COL03684, mRNA sequence.

VERSION AU106341

KEYWORDS AU106341.1 GI:13555862

SOURCE EST.

ORGANISM human.

REFERENCE 1 (bases 1 to 50)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.  
 Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
 Unpublished (2001)

JOURNAL Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: yusuzuki@ims.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

source

1. .50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="COL03684"

BASE COUNT 7 a 19 c 16 g 8 t

ORIGIN

Query Match 56.0%; Score 11.2; DB 107; Length 50;  
 Best Local Similarity 81.2%; Pred. No. 1.4e+05;  
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gtaccagggtgagctt 17

Db 11 GTACCGCGCTGAGCT 26  
 ||||| |||||

RESULT 15

R84653

LOCUS

DEFINITION

R84653 34 bp mRNA EST 14-AUG-1995  
 Y038c05.r1 Soares adult brain N2b4HB55Y Homo sapiens cDNA clone  
 IMAGE:180200 5' similar to gb:M92302 DIHYDROXYRINDINE-SENSITIVE

L-TYPE, BRAIN CALCIUM CHANNEL (HUMAN);, mRNA sequence.

ACCESSION R84653

VERSION R84653.1

KEYWORDS GI:943059

SOURCE EST.

ORGANISM human.

REFERENCE 1 (bases 1 to 34)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman  
 ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,  
 Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston  
 ,R., Williamson,A., Woldmann,P. and Wilson,R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 High quality sequence starts: 1  
 High quality sequence stops: 1  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Trace considered overall poor quality  
 Seq primer: M13RPI  
 High quality sequence stop: 1.

## FEATURES

source

1. .34  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3826844"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:180200"  
 /clone\_lib="Soares adult brain N2b4HB55Y"  
 /sex="Male"  
 /dev\_stage="55-year old"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /notes="Organ: brain; Vector: pT7T3D (Pharmacia) with a  
 modified polylinker; Site:1: Not 1; Site 2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTACCAATCTGAAGTGGAGCGCGCTTTTTTTTTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT7T3 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 53. Library constructed by Bento  
 Soares and M.Fatima Bonaldo. The adult brain RNA was  
 provided by Dr. Donald H. Gilden. Tissue was acquired  
 17-18 hours after death which occurred in consequence of a  
 ruptured aortic aneurysm. RNA was prepared from a pool of  
 tissues representing the following areas of the brain:  
 frontal, parietal, temporal and occipital cortex from the  
 left and right hemispheres, subcortical white matter,  
 basal ganglia, thalamus, cerebellum, midbrain, pons and  
 medulla."

BASE COUNT 11 a 11 c 6 g 5 t 1 others

ORIGIN

Query Match 55.0%; Score 11; DB 188; Length 34;  
 Best Local Similarity 70.0%; Pred. No. 1.7e+05;

Matches 14; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 agtaccaggtgagtcttag 20

Db 4 AGTCCCGAGGAAGNCTCAG 23

Search completed: October 2, 2001, 15:00:58  
Job time: 10827 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 2, 2001, 15:56:45 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-12  
Perfect score: 20  
Sequence: 1 cctgacatcagtagcatctc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_p11: \*  
13: gb\_p12: \*  
14: gb\_p13: \*  
15: gb\_p14: \*  
16: gb\_ba1: \*  
17: gb\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rod: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rod: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pl: \*  
48: em\_ro: \*  
49: em\_sts: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_vi: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_vil: \*  
59: gb\_vil2: \*  
60: gb\_htg1: \*  
61: gb\_htg2: \*  
62: gb\_htg3: \*  
63: gb\_htg4: \*  
64: gb\_htg5: \*  
65: gb\_htg6: \*  
66: gb\_htg7: \*  
67: gb\_htg8: \*  
68: gb\_htg9: \*  
69: gb\_htg10: \*  
70: gb\_htg11: \*  
71: gb\_htg12: \*  
72: gb\_htg13: \*  
73: gb\_htg14: \*  
74: gb\_htg15: \*  
75: gb\_htg16: \*  
76: gb\_htg17: \*  
77: gb\_htg18: \*  
78: gb\_htg19: \*  
79: gb\_htg20: \*  
80: gb\_htg21: \*  
81: gb\_htg22: \*  
82: gb\_htg23: \*  
83: gb\_htg24: \*  
84: gb\_htg25: \*  
85: gb\_pr1: \*  
86: gb\_pr2: \*  
87: gb\_pr3: \*  
88: gb\_pr4: \*  
89: gb\_pr5: \*  
90: gb\_pr6: \*  
91: gb\_pr7: \*  
92: gb\_pr8: \*  
93: gb\_pr9: \*  
94: gb\_rol: \*  
95: gb\_ro2: \*  
96: gb\_in4: \*  
97: gb\_pr10: \*  
98: em\_ba3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB ID	Description
C 1	13.8	69.0	17	9 AR046638	AR046638 Sequence
C 2	13.8	69.0	17	10 I53690	I53690 Sequence 14
C 3	13.8	69.0	21	10 AX095079	AX095079 Sequence
C 4	13.8	69.0	30	10 I27043	I27043 Sequence 64
C 5	13.6	68.0	42	9 AR021414	AR021414 Sequence
C 6	13.6	68.0	42	9 AR042976	AR042976 Sequence
C 7	13.6	68.0	42	10 I43923	I43923 Sequence 12
C 8	13.6	68.0	42	10 I62967	I62967 Sequence 64

```
c 9 13.6 68.0 42 10 188720 188720 Sequence 64
c 10 13.6 68.0 45 97 HUMIGHABH L06994 Human Ig re
c 11 13.2 66.0 30 9 AR063732 AR063732 Sequence 17
c 12 13.2 66.0 30 10 I30095 I30095 Sequence 17
c 13 13.2 66.0 23 9 AR020665 AR020665 Sequence 18
c 14 12.8 64.0 21 9 A51533 A51533 Sequence 18
c 15 12.8 64.0 21 9 A51561 A51561 Sequence 21
c 16 12.8 64.0 21 9 A87815 A87815 Sequence 11
c 17 12.8 64.0 21 9 A91147 A91147 Sequence 11
c 18 12.8 64.0 21 9 AR084287 AR084287 Sequence 11
c 19 12.8 64.0 21 10 I95565 I95565 Sequence 20
c 20 12.8 64.0 27 10 I25860 I25860 Sequence 20
c 21 12.8 64.0 27 10 I43497 I43497 Sequence 20
c 22 12.4 62.0 18 9 AR047426 AR047426 Sequence 11
c 23 12.4 62.0 18 10 I54478 I54478 Sequence 22
c 24 12.4 62.0 38 9 AR047211 AR047211 Sequence 22
c 25 12.4 62.0 38 9 AR057887 AR057887 Sequence 20
c 26 12.4 62.0 38 10 I54263 I54263 Sequence 20
c 27 12.4 62.0 41 9 AR050319 AR050319 Sequence 20
c 28 12.2 61.0 27 9 A22449 A22449 JH reporter
c 29 12.2 61.0 31 9 AR091207 AR091207 Sequence
c 30 12.2 61.0 40 9 AX036346 AX036346 Sequence
c 31 12.2 61.0 42 9 AR021372 AR021372 Sequence
c 32 12.2 61.0 42 9 AR042934 AR042934 Sequence
c 33 12.2 61.0 42 10 I62925 I62925 Sequence 20
c 34 12.2 61.0 42 10 I88678 I88678 Sequence 20
c 35 12.2 61.0 44 9 AR077278 AR077278 Sequence
c 36 12.2 61.0 48 10 I05886 I05886 Sequence 3
c 37 12.2 61.0 48 10 I05931 I05931 Sequence 3
c 38 12.2 61.0 48 10 I09165 I09165 Sequence 3
c 39 12 60.0 20 9 AR073966 AR073966 Sequence
c 40 12 60.0 20 9 AR105521 AR105521 Sequence
c 41 12 60.0 20 10 I27265 I27265 Sequence 35
c 42 12 60.0 24 9 A09906 A09906 Probe. 1/19
c 43 12 60.0 29 9 AX002747 AX002747 Sequence
c 44 12 60.0 29 9 AX003151 AX003151 Sequence
c 45 12 60.0 29 9 AX018567 AX018567 Sequence

ALIGNMENTS

RESULT 1
AR046638/c 17 bp DNA 29-SEP-1999
LOCUS AR046638 Sequence 1431 from patent US 5817796. PAT
DEFINITION AR046638
ACCESSION AR046638
VERSION AR046638.1 GI:5968103
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 17)
AUTHORS Stinchcomb,D.T., Draper,K., McSwiggen,J. and Jarvis,T.
TITLE C-myb ribozymes having 2'-5'-linked adenylyate residues
JOURNAL Patent: US 5817796-A 1431 06-OCT-1998;
FEATURES
Location/Qualifiers
1..17
BASE COUNT 2 a 4 c 5 g 6 t
ORIGIN
Query Match 69.0%; Score 13.8; DB 9; Length 17;
Best Local Similarity 88.2%; Pred. No. 9.4e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 4 gacatcagtagcatctc 20
Db 17 GACATCAGGAGCAACTC 1

RESULT 2
AR046638/c 21 bp DNA 30-MAR-2001
LOCUS AX095079 Sequence 257 from Patent WO0118250. PAT
DEFINITION AX095079
ACCESSION AX095079
VERSION AX095079.1 GI:13511282
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
1 (bases 1 to 21)
AUTHORS Lander,E.S., Gargill,M., Ireland,J.S., Bolk,S., Daley,G.Q. and
McCarthy,J.J.
TITLE Single nucleotide polymorphisms in genes
JOURNAL Patent: WO 0118250-A 257 15-MAR-2001;
WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH (US) ; Millennium
Pharmaceuticals, Inc. (US)
FEATURES
Location/Qualifiers
1..21
BASE COUNT 7 a 4 c 6 g 3 t 1 others
ORIGIN
Query Match 69.0%; Score 13.8; DB 10; Length 21;
Best Local Similarity 78.9%; Pred. No. 9.3e+03;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 cctgacatcagtagcatctc 19
Db 19 CTTGCCATYAGTGGCATCT 1

RESULT 4
I27043/c 30 bp DNA 06-FEB-1997
LOCUS I27043 Sequence 64 from patent US 5563036. PAT
DEFINITION I27043
ACCESSION I27043
VERSION I27043.1 GI:1817819
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
```

Query Match	Score 13.6;	DB 9;	Length 42;	LOCUS	100.00%	10-AUG-1998
	68.08;			I88720	42 bp	DNA
						PAT

```

DEFINITION Sequence 64 from patent US 5719032.
ACCESSION I88720
VERSION I88720.1 GI:3408660
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 42)
AUTHORS Vielkind,J.R.
TITLE Melanoma and prostate cancer specific antibodies for
immunodetection and immunotherapy
JOURNAL Patent: US 5719032-A 64 17-FEB-1998;
FEATURES Location/Qualifiers
source 1..42
BASE COUNT 5 a 8 c 15 g 14 t
ORIGIN

Query Match 68.0%; Score 13.6; DB 10; Length 42;
Best Local Similarity 80.0%; Pred. No. 1.2e+04;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cctgacatcagtagcatctc 20
||| ||||| ||||| |
Db 35 CCAGACATCAAAAGCATCAC 16

RESULT 10
HUMIGHABH/c
LOCUS HUMIGHABH 45 bp mRNA PRI 09-NOV-1994
DEFINITION Human Ig rearranged mu-chain mRNA V-region, partial cds (from clone
FBM-8).
ACCESSION L06994
VERSION L06994.1 GI:184940
KEYWORDS V-region; antigen binding region; complementarity determining
region; immunoglobulin heavy chain; mu-immunoglobulin; processed
gene.
SOURCE Homo sapiens 13 week fetus bone marrow cDNA to mRNA.
ORGANISM Homo sapiens

REFERENCE 1 (sites)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Ichihara,Y., Matsuoka,H. and Kurosawa,Y.
Organization of human immunoglobulin heavy chain diversity gene
loci
JOURNAL EMBO J. 7 (13), 4141-4150 (1988)
MEDLINE 89210798
AUTHORS 2 (bases 1 to 45)
Raaphorst,F.M., Sanz,I., Vossen,J.M., Schuurman,R.K.B. and van
Tol,M.J.D.
TITLE Immunoglobulin heavy chain CDR3 regions in human fetal organs:
Evidence for selection of rearrangements on the basis of CDR3
length
JOURNAL Unpublished (1993)
FEATURES Location/Qualifiers
source 1..45
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="B lymphocyte"
/dev_stage="13 week fetus"
/tissue_type="bone marrow"
/map="14q32.33"
1..45
/gene="IGHM"
<1..>45
/gene="IGHM"
/note="G00-120-086"
BASE COUNT 7 a 5 c 18 g 15 t
ORIGIN

gene
V_region

Query Match 68.0%; Score 13.6; DB 97; Length 45;
Best Local Similarity 80.0%; Pred. No. 1.2e+04;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cctgacatcagtagcatctc 20
||| ||||| ||||| |
Db 35 CCAGACATCAAAAGCATCAC 16

RESULT 11
AR063732
LOCUS AR063732 30 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 17 from patent US 5846720.
ACCESSION AR063732
VERSION AR063732.1 GI:5993040
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 30)
AUTHORS Foulkes,J.Gordon, Liechtfried,F.E., Pieler,C., Stephenson,J.R. and
Case,C.C.
TITLE Methods of determining chemicals that modulate expression of genes
associated with cardiovascular disease
JOURNAL Patent: US 5846720-A 17 08-DEC-1998;
FEATURES Location/Qualifiers
source 1..30
/organism="unknown"
BASE COUNT 11 a 7 c 5 g 7 t
ORIGIN

Query Match 66.0%; Score 13.2; DB 9; Length 30;
Best Local Similarity 83.3%; Pred. No. 2e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ctgacatcagtagcatct 19
||||| ||||| |||||
Db 6 CTGAATCAGCAGCACCT 23

RESULT 12
I30095
LOCUS I30095 30 bp DNA PAT 06-FEB-1997
DEFINITION Sequence 17 from patent US 5580722.
ACCESSION I30095
VERSION I30095.1 GI:1820886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 30)
AUTHORS Foulkes,J.Gordon, Liechtfried,F.E., Pieler,C., Stephenson,J.R. and
Case,C.C.
TITLE Methods of determining chemicals that modulate transcriptionally
expression of genes associated with cardiovascular disease
JOURNAL Patent: US 5580722-A 17 03-DEC-1996;
FEATURES Location/Qualifiers
source 1..30
/organism="unknown"
BASE COUNT 11 a 7 c 5 g 7 t
ORIGIN

Query Match 66.0%; Score 13.2; DB 10; Length 30;
Best Local Similarity 83.3%; Pred. No. 2e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ctgacatcagtagcatct 19
||||| ||||| |||||
Db 6 CTGAATCAGCAGCACCT 23

RESULT 13

```

```
AR020665/c
LOCUS       AR020665      23 bp      DNA
DEFINITION   Sequence 59 from patent US 5789184.
ACCESSION   AR020665
VERSION     AR020665.1  GI:3975280
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 23)
AUTHORS     Fowles,D.M., Broach,J., Manfredi,J., Klein,C., Murphy,A.J.,
            Paul,J. and Trueheart,J.
TITLE       Yeast cells engineered to produce pheromone system protein
            surrogates, and uses therefor
JOURNAL     Patent: US 5789184-A 59 04-AUG-1998;
FEATURES
            Location/Qualifiers
            1..23
            /organism="unknown"
BASE COUNT   4 a      8 c      5 g      6 t
ORIGIN
            65.0%; Score 13; DB 9; Length 23;
            Best Local Similarity 100.0%; Pred. NO. 2.6e+04;
            Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  7 atcagtagcatct 19
    |||||
Db  18 ATCAGTAGCATCT 6

RESULT 14
A51533
LOCUS       A51533      21 bp      DNA
DEFINITION   Sequence 18 from Patent EP0728842.
ACCESSION   A51533
VERSION     A51533.1  GI:2304354
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 21)
AUTHORS     Audonnet,J.F., Bublout,M.J., Dartell,R.J., Duinat,C.V., Laplace,E.L.
            and Riviere,M.A.
TITLE       Live recombinant avian vaccine based on an avianherpes virus,
            against Gumboro disease
JOURNAL     Patent: EP 0728842-A 18 28-AUG-1996;
            RHONE MERIEUX (FR)
COMMENT     Other publication FR 2728794 960705
            Other publication CA 2166371 960701
            Other publication AU 4063095 960711.
FEATURES
            Location/Qualifiers
            1..21
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT   5 a      7 c      4 g      5 t
ORIGIN

Query Match      64.0%; Score 12.8; DB 9; Length 21;
Best Local Similarity 87.5%; Pred. NO. 3.3e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  3 tgacatcagtagcatc 18
    |||||
Db  2 TGACATCACTAGCGTC 17

RESULT 15
A51561
LOCUS       A51561      21 bp      DNA
DEFINITION   Sequence 21 from Patent EP0719864.
ACCESSION   A51561
```

```
VERSION     A51561.1  GI:2304389
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 21)
AUTHORS     Audonnet,J.F., Bublout,M.J., Dartell,R.J., Duinat,C.V., Laplace,E.L.
            and Riviere,M.A.
TITLE       Recombinant live avian vaccin, using an avian herpes virus as
            vector
JOURNAL     Patent: EP 0719864-A 21 03-JUL-1996;
            RHONE MERIEUX (FR)
COMMENT     Other publication FR 2728795 960705
            Other publication CA 2166367 960701
            Other publication AU 4071595 960711.
FEATURES
            Location/Qualifiers
            1..21
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT   5 a      7 c      4 g      5 t
ORIGIN

Query Match      64.0%; Score 12.8; DB 9; Length 21;
Best Local Similarity 87.5%; Pred. NO. 3.3e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  3 tgacatcagtagcatc 18
    |||||
Db  2 TGACATCACTAGCGTC 17

Search completed: October 2, 2001, 15:56:47
Job time: 14171 sec
```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:00:58 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-12  
Perfect score: 20  
Sequence: 1 cctgacatcagtagcatctc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues  
Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

em\_esthum10:\*  
44: em\_esthum11:\*  
45: em\_esthum12:\*  
46: em\_esthum13:\*  
47: em\_esthum14:\*  
48: em\_esthum15:\*  
49: em\_esthum16:\*  
50: em\_esthum17:\*  
51: em\_esthum18:\*  
52: em\_esthum19:\*  
53: em\_esthum20:\*  
54: em\_esthum21:\*  
55: em\_esthum22:\*  
56: em\_esthum23:\*  
57: em\_esthum24:\*  
58: em\_esthum25:\*  
59: em\_esthum26:\*  
60: em\_esthum27:\*  
61: em\_esthum28:\*  
62: em\_estin1:\*  
63: em\_estin2:\*  
64: em\_estin3:\*  
65: em\_estin4:\*  
66: em\_estin5:\*  
67: em\_estom1:\*  
68: em\_estom2:\*  
69: em\_estov1:\*  
70: em\_estov2:\*  
71: em\_estpl1:\*  
72: em\_estpl2:\*  
73: em\_estpl3:\*  
74: em\_estpl4:\*  
75: em\_estpl5:\*  
76: em\_estpl6:\*  
77: em\_estpl7:\*  
78: em\_estpl8:\*  
79: em\_estpl9:\*  
80: em\_estpl10:\*  
81: em\_estro1:\*  
82: em\_estro2:\*  
83: em\_estro3:\*  
84: em\_estro4:\*  
85: em\_estro5:\*  
86: em\_estro6:\*  
87: em\_estro7:\*  
88: em\_estro8:\*  
89: em\_estro9:\*  
90: em\_estro10:\*  
91: em\_estro11:\*  
92: em\_estro12:\*  
93: em\_estro13:\*  
94: em\_estro14:\*  
95: em\_estro15:\*  
96: em\_estro16:\*  
97: em\_estro17:\*  
98: em\_estro18:\*  
99: em\_estro19:\*  
100: em\_estro20:\*  
101: gb\_est25:\*  
102: gb\_est26:\*  
103: gb\_est27:\*  
104: gb\_est28:\*  
105: gb\_est29:\*  
106: gb\_est30:\*  
107: gb\_est31:\*  
108: gb\_est32:\*  
109: gb\_est41:\*  
110: gb\_est42:\*  
111: gb\_est43:\*  
112: gb\_est44:\*  
113: gb\_est45:\*  
114: gb\_est46:\*  
115: gb\_est47:\*  
116:

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est110:\*  
191: gb\_est111:\*  
192: gb\_est112:\*  
193: gb\_est113:\*  
194: gb\_est114:\*  
195: gb\_est115:\*  
196: gb\_est116:\*  
197: gb\_est117:\*  
198: gb\_est118:\*  
199: gb\_est119:\*  
200: gb\_est120:\*  
201: gb\_est121:\*  
202: gb\_est122:\*  
203: gb\_est123:\*  
204: gb\_est124:\*  
205: gb\_est125:\*  
206: gb\_est126:\*  
207: gb\_est127:\*  
208: gb\_est128:\*  
209: gb\_est129:\*  
210: gb\_est130:\*  
211: gb\_est131:\*  
212: gb\_est132:\*  
213: gb\_est133:\*  
214: gb\_est134:\*  
215: gb\_est135:\*  
216: gb\_est136:\*  
217: gb\_est137:\*  
218: gb\_est138:\*  
219: gb\_est139:\*  
220: gb\_est140:\*  
221: gb\_est141:\*  
222: gb\_est142:\*  
223: gb\_est143:\*  
224: gb\_est144:\*  
225: gb\_est145:\*  
226: gb\_est146:\*  
227: gb\_est147:\*  
228: gb\_est148:\*  
229: gb\_est149:\*  
230: gb\_est150:\*  
231: gb\_est151:\*  
232: gb\_est152:\*  
233: gb\_est153:\*  
234: gb\_est154:\*  
235: gb\_est155:\*  
236: gb\_est156:\*  
237: gb\_est157:\*  
238: gb\_est158:\*  
239: gb\_est159:\*  
240: gb\_est160:\*  
241: gb\_est161:\*  
242: gb\_est162:\*  
243: gb\_est163:\*  
244: gb\_est164:\*  
245: gb\_est165:\*  
246: gb\_est166:\*  
247: gb\_est167:\*  
248: gb\_est168:\*  
249: gb\_est169:\*  
250: gb\_est170:\*  
251: gb\_est171:\*  
252: gb\_est172:\*  
253: gb\_est173:\*  
254: gb\_est174:\*  
255: gb\_est175:\*  
256: gb\_est176:\*  
257: gb\_est177:\*  
258: gb\_est178:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



SUMMARIES									
Result No.	Score	Query Match	Length	DB	ID	Description			
-----									
C	1	13.2	66.0	19	250	AZ848554	AZ848554	2M0149A07	
	2	12.6	63.0	49	16	AI098615	AI098615	ue31h01.y	
	3	12.4	62.0	31	13	AA910858	AA910858	OK85C01.s	
	4	12	60.0	40	104	AI962543	AI962543	wq53q05.x	
	5	12	60.0	49	14	AA975152	AA975152	Om99c07.s	
	6	11.8	59.0	34	23	AI657872	AI657872	fc14f02.y	
	7	11.8	59.0	41	258	HSMC04C05	X8541.H.sapiens.D		
C	8	11.8	59.0	46	249	AZ758330	AZ758330	IM0550115	
C	9	11.8	59.0	49	24	AI67413	AI67413	wi95g10.x	
	10	11.6	58.0	49	2	AA087870	AA087870	mn97all.r	
	11	11.6	58.0	49	104	AI949145	AI949145	wq08b11.x	
	12	11.4	57.0	47	158	H55380	H55380	CHR220319.C	
	13	11.2	56.0	27	249	AZ769505	AZ769505	IM0570L17	
	14	11.2	56.0	31	17	AI174159	AI174159	v284e05.r	
	15	11.2	56.0	31	246	AZ620204	AZ620204	IM0452F14	
C	16	11.2	56.0	31	249	AZ777633	AZ777633	2M0012K21	
	17	11.2	56.0	42	244	AZ456886	AZ456886	IM0260G01	
	18	11.2	56.0	45	244	AZ465721	AZ465721	IM0275O14	
	19	11.2	56.0	49	6	AA388129	AA388129	vc86f05.r	
C	20	11	55.0	34	17	AI168400	AI168400	qa25e03.s	
C	21	11	55.0	35	153	BG393308	BG393308	602411521	
C	22	11	55.0	37	14	AA95811	AA95811	os05e12.s	
	23	11	55.0	37	17	AI208016	AI208016	gg60f11.x	
	24	11	55.0	43	190	W62348	W62348	md39c05.r1	
C	25	11	55.0	50	107	AU103398	AU103398	AJ239886	
	26	11	55.0	50	107	AU104027	AU104027	AJ239886	
C	27	11	55.0	50	107	AU107009	AU107009	AL451432.T.brucel	
C	28	10.8	54.0	22	244	AZ459715	AZ459715	IM0264E13	
C	29	10.8	54.0	26	250	AZ814137	AZ814137	2M0081B16	
C	30	10.8	54.0	34	250	AZ856611	AZ856611	2M0161P06	
C	31	10.8	54.0	39	104	AJ239886	AJ239886	AJ239886	
	32	10.8	54.0	43	246	AZ592784	AZ592784	IM0403O24	
C	33	10.8	54.0	43	258	TA5F100	TA5F100		
	34	10.8	54.0	46	189	T53599	T53599	ya97f09.sl	
	35	10.6	53.0	28	17	AI187132	AI187132	qa42f11.x	
	36	10.6	53.0	30	242	AZ345667	AZ345667	IM0080O22	
	37	10.6	53.0	34	17	AI188358	AI188358	qq08c03.x	
	38	10.6	53.0	35	242	AZ389531	AZ389531	IM0150P07	
C	39	10.6	53.0	36	247	AZ659063	AZ659063	IM0536N10	
C	40	10.6	53.0	39	191	Z20226	Z20226	HSAAABOLR.S	
C	41	10.6	53.0	40	14	AA999748	AA999748	os54g04.s	
C	42	10.6	53.0	40	106	AU013065	AU013065	AU013065	
C	43	10.6	53.0	40	106	AU013075	AU013075		
C	44	10.6	53.0	40	139	BE739544	BE739544	601556A85	
	45	10.6	53.0	40	244	AZ462628	AZ462628	IM0269E08	
-----									
ALIGNMENTS									
-----									
RESULT 1									
AZ848554/C									
LOCUS									
DEFINITION	2M0149A07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0149A07 R, DNA sequence.								
ACCESSION	AZ848554								
VERSION	GSS								
KEYWORDS	house mouse.								
SOURCE	Mus musculus								
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 19)								
REFERENCE	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.								
AUTHORS									
-----									
Query Match	66.0%; Score 13.2; DB 250; Length 19;								
Best Local Similarity	83.3%; Pred. No. 1.7e+04;								
Matches	15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;								
QY	1 cctgcacatcagtagcatc 18								
Db	19 CCTGACGACCTAGCAGC 2								
RESULT 2									
AI098615									
LOCUS									
DEFINITION	AI098615 49 bp mRNA EST 20-AUG-1998 ue31h01.y1 Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:1482001 5' similar to TR:P97298 P97298 STROMAL CELL DERIVED FACTOR 3 ;, mRNA sequence.								
ACCESSION	AI098615								
VERSION	AI098615.1 GI:3448140								
KEYWORDS	EST.								
SOURCE	house mouse.								
ORGANISM	Mus musculus								
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 49)								
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,								

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The Rashi-HMI Mouse EST Project  
Unpublished (1996)

# TITLE JOURNAL COMMENT

Contact: Marra M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:930357

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: custom primer used

High quality sequence stop: 1.

## FEATURES source

```
1. .49
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1482001"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
(ATGTGCGCTTTTCTTTTCTTTT); double-stranded cDNA was
ligated to a DraIII adaptor [TCTGGCCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection should
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTGCTCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGACACA."
```

BASE COUNT 13 a 18 c 10 g 8 t

## ORIGIN

```
Query Match 63.0%; Score 12.6; DB 16; Length 49;
Best Local Similarity 78.9%; Pred. No. 3.8e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 cctgacatcagtagcatct 19
||||||| |||||
Db 23 CCTGACATCCACGACCT 41
```

## RESULT 3

```
AA910858 31 bp mRNA EST 13-APR-1998
LOCUS OK85c01.s1 NCI_CGAP_kid3 Homo sapiens cDNA clone IMAGE:1520736 3'
DEFINITION similar to TR:Q13531 Q13531 DISHEVELLED ;, mRNA sequence.
```

```
ACCESSION AA910858
VERSION AA910858.1 GI:3050148
KEYWORDS EST.
```

## SOURCE

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

## REFERENCE

```
1 (bases 1 to 31)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

## AUTHORS

```
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
```

Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Cloning Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 1.

## FEATURES source

```
1. .31
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1520736"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."
```

BASE COUNT 4 a 6 c 16 g 5 t

## ORIGIN

```
Query Match 62.0%; Score 12.4; DB 13; Length 31;
Best Local Similarity 92.9%; Pred. No. 4.5e+04;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 4 gacatcagtagcatc 17

Db 9 GGCATCAGTAGCAT 22

## RESULT 4

```
AI962543 40 bp mRNA EST 08-MAR-2000
LOCUS wg53q05.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2475032 3'
DEFINITION similar to TR:O88247 O88247 MSZF52 ;, mRNA sequence.
```

```
ACCESSION AI962543
VERSION AI962543.1 GI:5755256
KEYWORDS EST.
```

## SOURCE

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

## REFERENCE

```
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

## TITLE

```
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
```

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.

## SOURCE

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

## REFERENCE

```
1 (bases 1 to 40)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

## AUTHORS

```
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
```

Trace considered overall poor quality

Insert Length: 884 Std Error: 0.00

Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 40

#### FEATURES

source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2475032"  
/clone\_lib="NCI\_CGAP\_GC6"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."  
9 a 13 c 3 g 15 t

BASE COUNT  
ORIGIN

Query Match 60.0%; Score 12; DB 104; Length 40;  
Best Local Similarity 75.0%; Pred. No. 7.4e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 cctgacatcagtagcatctc 20  
||| |||| ||||| ||

Db 19 CCTCACATTCATGACATTTC 38

RESULT 5

AA975152

LOCUS

DEFINITION AA975152 49 bp mRNA EST 22-MAY-1998  
om99c07.s1 NCI\_CGAP\_Kid3 Homo sapiens cDNA clone IMAGE:1555308 3'  
similar to TR:O19048 O19048 HNRNP-E1 PROTEIN. ;, mRNA sequence.

ACCESSION AA975152

VERSION AA975152.1 GI:3150944

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 1.

FEATURES

source

1. 49  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1555308"  
/clone\_lib="NCI\_CGAP\_Kid3"  
/lab\_host="DH10B"

/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dw) primer,

double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT7T3 vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo.

BASE COUNT 13 a 8 c 11 g 17 t  
ORIGIN

Query Match 60.0%; Score 12; DB 14; Length 49;  
Best Local Similarity 75.0%; Pred. No. 7.6e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 cctgacatcagtagcatctc 20

||| ||||| ||| |||||

Db 21 CCAGACATCTGGCGGATCTC 40

RESULT 6

AI657872

LOCUS

DEFINITION AI657872 34 bp mRNA EST 06-MAY-1999  
fc14f02.y1 zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to  
SW:T2D6\_HUMAN Q15545 TRANSCRIPTION INITIATION FACTOR TFIID 55 KD  
SUBUNIT ;, mRNA sequence.

ACCESSION AI657872

VERSION AI657872.1 GI:47555540

KEYWORDS EST.

SOURCE zebrafish.

ORGANISM

Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Rasbora; Danio.

REFERENCE

AUTHORS

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy  
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood  
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,  
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,  
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.  
and Wilson, R.

WashU zebrafish EST Project 1998

UNPUBLISHED (1998)

Other ESTs: fc14f02.x1

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: [zbrafish@watson.wustl.edu](mailto:zbrafish@watson.wustl.edu)

CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clone distribution: Genome Systems, St. Louis,

Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:

[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama

(web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and

ReSourceZentrumPrimarDatenbank, Berlin, Germany (web address:

[www.rzpd.de](http://www.rzpd.de))

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: T3 ET from Amersham

High quality sequence stop: 1.

FEATURES

source

1. 34

Location/Qualifiers

/organism="Danio rerio"

/db\_xref="taxon:7955"

/clone\_lib="Zebrafish WashU MPIMG EST"

/sex="mixed"

/tissue\_type="26 somite embryos, adult livers, shield

stage embryos"

/lab\_host="XLI-blue MRF"

/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st

strand cDNA was primed with a Not I - oligo(dw) primer

(5'pCAGTAGTCTAGATCGGCGGCCCTTTTCTTTTCTTTTCTTTTCTTTT3')

double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 13 a 7 c 11 g 3 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 23; Length 34;  
Best Local Similarity 86.7%; Pred. No. 9.1e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gacatcagtagc 18  
||||||| |||||  
Db 2 GACATCAGCAGC 16

RESULT 7  
HSMC04C05 HSMC04C05 41 bp DNA GSS 29-MAY-1997  
LOCUS H.sapiens DNA for trapped exon (ID RMC04C05), genomic survey  
DEFINITION sequence.  
X88541  
X88541.1 GI:1437552  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 41)  
AUTHORS Chen,H.M., Rossier,C., Chrast,R. and Antonarakis,S.E.  
TITLE Cloning of trapped exons from human chromosome 21  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 41)  
AUTHORS Antonarakis,S.E.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAR-1995) Stylianos E. Antonarakis, Division of Medical Genetics, University and Cantonal Hospital of Geneva, CMU, 1 rue Michel-Servet, 1211 Geneva, SWITZERLAND  
REFERENCE 3 (bases 1 to 41)  
AUTHORS Chen,H., Chrast,R., Rossier,C., Morris,M.A., Lalioti,M.D. and Antonarakis,S.E.  
TITLE Cloning of 559 potential exons of genes of human chromosome 21 by exon trapping  
JOURNAL Genome Res. 6 (8), 747-760 (1996)  
MEDLINE 97011340

FEATURES  
Location/Qualifiers  
1..41  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
1..41  
/note="trapped exon"  
BASE COUNT 12 a 8 c 6 g 12 t 3 others  
ORIGIN

Query Match 59.0%; Score 11.8; DB 258; Length 41;  
Best Local Similarity 86.7%; Pred. No. 9.4e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 catcagtagc 20

Db 14 CAACAGAGCATCTC 28  
|| ||| ||||| |||||

RESULT 8  
AZ758330/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AZ758330 46 bp DNA GSS 16-FEB-2001  
1M0550115F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0550115 F, DNA sequence.

AZ758330  
AZ758330.1 GI:12864016  
GSS.  
house mouse.  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 46)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0550 row: 1 column: 15  
Seq primer: CGTGTAAACGACGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 46.

FEATURES  
Location/Qualifiers

1..46  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0550115"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, P-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gil4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 12 a 6 c 11 g 17 t  
ORIGIN

Query Match 59.0%; Score 11.8; DB 249; Length 46;  
Best Local Similarity 86.7%; Pred. No. 9.5e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 4 gacatcagtagcatc 18
Db 18 GACATCAGTAGAACC 4

RESULT 9
A1767413/c
LOCUS
DEFINITION A1767413 49 bp mRNA EST 21-DEC-1999
similar to WP:F55F8.5 Cell1196 ;, mRNA sequence.
IMAGE:2401122 3'
ACCESSION A1767413
VERSION A1767413.1 GI:5233842
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 49)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 969 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. 49
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2401122"
/clone_lib="NCI-CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 12 a 11 c 12 g 14 t
ORIGIN

Query Match 59.0%; Score 11.8; DB 24; Length 49;
Best Local Similarity 86.7%; Pred. No. 9.6e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tgacatcagtagcatc 17
Db 36 TGACATCAGTAAAT 22

RESULT 10
AA087870
LOCUS
DEFINITION AA087870 49 bp mRNA EST 15-FEB-1997
mn97all.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
IMAGE:551996 5' similar to TR:G435967 G435967 APOLIPOPROTEIN F. ;,

```

```

mRNA sequence.
ACCESSION AA087870
VERSION AA087870.1 GI:1631078
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 49)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Scheilenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:332788
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 revl ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers
1. 49
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:551996"
/clone_lib="Stratagene mouse lung 937302"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 month old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dt. 6-8 month old female lung and 1.5 year old male lung
were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP
XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3'
adaptor sequence: 5' CTCAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 10 a 17 c 8 g 14 t
ORIGIN

Query Match 58.0%; Score 11.6; DB 2; Length 49;
Best Local Similarity 77.8%; Pred. No. 1.2e+05;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cctgacatcagtagcatc 18
Db 2 CCTGACATGCATAGCCTC 19

RESULT 11
A1949145
LOCUS
DEFINITION A1949145 49 bp mRNA EST 08-MAR-2000
wq08b11.x1 NCI-CGAP_Kid12 Homo sapiens cDNA clone
similar to TR:O60455 O60455 POLYADENYLATE BINDING
PROTEIN-INTERACTING PROTEIN-1. ;, mRNA sequence.
ACCESSION A1949145
VERSION A1949145.1 GI:5741455
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 49)

```

**AUTHORS** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
**JOURNAL** Tumor Gene Index  
**COMMENT** Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-f@mail.nih.gov](mailto:cgaps-f@mail.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
 Insert Length: 779 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 1.

#### FEATURES

source  
 1. .49  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2470653"  
 /clone\_lib="NCI\_CGAP\_Kid12"  
 /tissue\_type="2 pooled tumors (clear cell type)"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Plasmid DNA from the normalized library NCI\_CGAP\_Kid5 was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (cloneIDs 1323912-1325831, 1471368-1472903 and  
 1492104-1493255). Subtraction by Bento Soares and M.  
 Fatima Bonaudo."

**BASE COUNT** 10 a 13 c 9 g 16 t 1 others  
**ORIGIN**

Query Match 58.0%; Score 11.6; DB 104; Length 49;  
 Best Local Similarity 77.8%; Pred. No. 1.2e+05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

**QY** 2 ctgacatcagtagcatct 19  
 |||||  
**Db** 18 CTTACTGCAGTCGCATCT 35

**RESULT** 12  
**H55380**  
**LOCUS** H55380 47 bp mRNA EST 07-DEC-1995  
**DEFINITION** CHR220319 Chromosome 22 exon Homo sapiens cDNA clone C22\_405 5',  
 mRNA sequence.

**ACCESSION** H55380  
**VERSION** H55380.1 GI:1108246  
**KEYWORDS** EST.  
**SOURCE** human.

**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 47)  
**AUTHORS** Trofatter, J.A., Long, K.R., Murrell, J.R., Stotler, C.J., Gusella, J.F.  
 and Buckler, A.J.

**TITLE** An expression-independent catalog of genes from human chromosome 22  
**JOURNAL** Genome Res. 5 (3), 214-224 (1995)  
**MEDLINE** 96159527  
**COMMENT** Contact: Buckler AJ  
 Molecular Neurogenetics Unit  
 Massachusetts General Hospital  
 Building 149, 13th St., Charlestown MA 02129  
 Tel: 6177249616

Fax: 6177265736  
 Email: [buckler@helix.mgh.harvard.edu](mailto:buckler@helix.mgh.harvard.edu)  
 Seq primer: T3.

#### FEATURES

Location/Qualifiers  
 source  
 1. .47  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="C22\_405"  
 /clone\_lib="Chromosome 22 exon"  
 /lab\_host="E. coli DH5a"  
 /note="Vector: pBluescriptIIKS+; Site\_1: Sal I; Site\_2:  
 Bam HI (destroyed); Exons were isolated from human  
 chromosome 22 specific cosmids using a modification of  
 the method of exon amplification (Proc. Natl. Acad. Sci.  
 USA 88:4005-4009, 1991). Amplified exons were digested  
 with Sal I and Bgl II and subsequently cloned into  
 pBluescriptIIKS+ at the Sal I and Bam HI sites."

**BASE COUNT** 15 a 11 c 12 g 9 t  
**ORIGIN**

Query Match 57.0%; Score 11.4; DB 158; Length 47;  
 Best Local Similarity 92.3%; Pred. No. 1.5e+05;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 7 atcagtagcatct 19  
 |||||  
**Db** 31 ATCATTAGCATCT 43

**RESULT** 13  
**AZ769505**

**LOCUS** AZ769505 27 bp DNA GSS 16-FEB-2001  
**DEFINITION** 1M057017F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M057017 F, DNA sequence.

**ACCESSION** AZ769505  
**VERSION** AZ769505.1 GI:12889709  
**KEYWORDS** GSS.  
**SOURCE** house mouse.

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
**REFERENCE** 1 (bases 1 to 27)  
**AUTHORS** Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.

**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Robert B. Weiss

University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0570 row: L column: 17  
 Seq primer: CCGTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 27.

#### FEATURES

Location/Qualifiers  
 source  
 1. .27  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M057017"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, P-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M."

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g114732114|gblAF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

BASE COUNT      9 a      3 c      8 g      7 t
ORIGIN
Query Match      56.0%; Score 11.2; DB 249; Length 27;
Best Local Similarity 81.2%; Pred. No. 1.8e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ctgacatcagtagcat 17
    ||| ||||| ||| |||
Db 5 CTGTATCATAGTAGAT 20

```

```

RESULT 14
A1174159
LOCUS      31 bp      mRNA      EST      07-OCT-1998
DEFINITION v284e05.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:1333184 5' similar to SW:NUIM_MOUSE P03888 NADH-UBIQUINONE
OXIDOREDUCTASE CHAIN 1 ; mRNA sequence.
A1174159
A1174159.1 GI:3720301
EST.
SOURCE      house mouse.
ORGANISM    Mus musculus

```

```

REFERENCE
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
University of Utah
Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0452 row: P column: 14
Seq primer: CACACAGGAACAGCATGACC
Class: plasmid ends
High quality sequence stop: 1.
Location/Qualifiers
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1333184"
/clone_lib="Soares_mammary_gland_NbMMG"

```

```

FEATURES
source
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1333184"
/clone_lib="Soares_mammary_gland_NbMMG"

```

```

FEATURES
source
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1333184"
/clone_lib="Soares_mammary_gland_NbMMG"

```

```

/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dr)
primer [5'
TGTTCACCAATCTGAAGTGGAGCGCGCAATGCTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

```

BASE COUNT      9 a      8 c      7 g      7 t
ORIGIN
Query Match      56.0%; Score 11.2; DB 17; Length 31;
Best Local Similarity 81.2%; Pred. No. 1.8e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cctgacatcagtagca 16
    ||| |||| |||| |
Db 13 CCTAACATTAGTAGAA 28

```

```

RESULT 15
A2620204
LOCUS      31 bp      DNA      GSS      13-DEC-2000
DEFINITION IM0452P14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0452P14 R, DNA sequence.
ACCESSION  A2620204
VERSION     A2620204.1 GI:11742394
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus

```

```

REFERENCE
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
University of Utah
Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0452 row: P column: 14
Seq primer: CACACAGGAACAGCATGACC
Class: plasmid ends
High quality sequence stop: 31.
Location/Qualifiers
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0452P14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource

```

```

FEATURES
source
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0452P14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource

```

```

FEATURES
source
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0452P14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource

```

```

FEATURES
source
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0452P14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource

```

(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWp42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 4 a 5 c 16 g 6 t  
ORIGIN

Query Match 56.0%; Score 11.2; DB 246; Length 31;  
Best Local Similarity 81.2%; Pred. No. 1.8e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ctgacatcagtagcat 17  
||||| |||||  
Db 3 CTGACAGCAGTGGCTT 18

Search completed: October 2, 2001, 15:01:00  
Job time: 10829 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:47 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-14

Perfect score: 20

Sequence: 1 ggttaggatgtgcgcgtca 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_ba3.\*

4: gb\_in1.\*

5: gb\_in2.\*

6: gb\_in3.\*

7: gb\_om.\*

8: gb\_ov.\*

9: gb\_pat1.\*

10: gb\_pat2.\*

11: gb\_ph.\*

12: gb\_pl1.\*

13: gb\_pl2.\*

14: gb\_pl3.\*

15: gb\_pl4.\*

16: em\_ba1.\*

17: em\_ba2.\*

18: em\_fun.\*

19: em\_htgo\_hum.\*

20: em\_htgo\_inv.\*

21: em\_htgo\_rod.\*

22: em\_htg\_hum1.\*

23: em\_htg\_hum2.\*

24: em\_htg\_hum3.\*

25: em\_htg\_hum4.\*

26: em\_htg\_hum5.\*

27: em\_htg\_hum6.\*

28: em\_htg\_hum7.\*

29: em\_htg\_hum8.\*

30: em\_htg\_inv1.\*

31: em\_htg\_inv2.\*

32: em\_htg\_other.\*

33: em\_htg\_rod.\*

34: em\_hum1.\*

35: em\_hum2.\*

36: em\_hum3.\*

37: em\_hum4.\*

38: em\_hum5.\*

39: em\_hum6.\*

40: em\_hum7.\*

41: em\_in.\*

42: em\_om.\*

43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_vil2.\*  
61: gb\_vil2.\*  
62: gb\_vil2.\*  
63: gb\_vil2.\*  
64: gb\_vil2.\*  
65: gb\_vil2.\*  
66: gb\_vil2.\*  
67: gb\_vil2.\*  
68: gb\_vil2.\*  
69: gb\_vil2.\*  
70: gb\_vil2.\*  
71: gb\_vil2.\*  
72: gb\_vil2.\*  
73: gb\_vil2.\*  
74: gb\_vil2.\*  
75: gb\_vil2.\*  
76: gb\_vil2.\*  
77: gb\_vil2.\*  
78: gb\_vil2.\*  
79: gb\_vil2.\*  
80: gb\_vil2.\*  
81: gb\_vil2.\*  
82: gb\_vil2.\*  
83: gb\_vil2.\*  
84: gb\_vil2.\*  
85: gb\_vil2.\*  
86: gb\_vil2.\*  
87: gb\_vil2.\*  
88: gb\_vil2.\*  
89: gb\_vil2.\*  
90: gb\_vil2.\*  
91: gb\_vil2.\*  
92: gb\_vil2.\*  
93: gb\_vil2.\*  
94: gb\_vil2.\*  
95: gb\_vil2.\*  
96: gb\_vil2.\*  
97: gb\_vil2.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	13.8	69.0	24	9 AR016339	AR016339 Sequence 18
C 2	13.8	69.0	24	10 I59963	I59963 Sequence 18
C 3	13.8	69.0	24	10 I86044	I86044 Sequence 18
C 4	13.8	69.0	30	10 I24249	I24249 Sequence 36
C 5	12.2	61.0	34	9 AX001078	AX001078 Sequence
6	12	60.0	26	9 AR087980	AR087980 Sequence 87
7	12	60.0	26	10 I80198	I80198 Sequence 15
8	12	60.0	32	10 I07710	I07710 Sequence 15

9 12 60.0 39 9 A92270  
c 10 12 60.0 42 95 S77075  
c 11 12 60.0 45 97 S63845  
c 12 11.8 59.0 20 10 E08846  
c 13 11.8 59.0 26 9 AR087898  
c 14 11.8 59.0 26 9 AR087920  
c 15 11.8 59.0 26 9 AR087948  
c 16 11.8 59.0 26 9 AR087958  
c 17 11.8 59.0 26 9 AR087960  
c 18 11.8 59.0 26 9 AR087974  
c 19 11.8 59.0 26 9 AR087976  
c 20 11.8 59.0 26 9 AR087978  
c 21 11.8 59.0 26 9 AR090813  
c 22 11.8 59.0 26 10 I80116  
c 23 11.8 59.0 26 10 I80138  
c 24 11.8 59.0 26 10 I80166  
c 25 11.8 59.0 26 10 I80176  
c 26 11.8 59.0 26 10 I80178  
c 27 11.8 59.0 26 10 I80192  
c 28 11.8 59.0 26 10 I80194  
c 29 11.8 59.0 26 10 I80196  
c 30 11.8 59.0 30 10 I24263  
c 31 11.8 59.0 35 9 A69153  
c 32 11.8 59.0 40 97 HSZ74584  
c 33 11.8 59.0 46 9 AX059019  
c 34 11.8 59.0 50 10 E16496  
c 35 11.6 58.0 23 9 AX003306  
c 36 11.6 58.0 24 9 AR083193  
c 37 11.6 58.0 25 9 AR090589  
c 38 11.6 58.0 26 9 AR090376  
c 39 11.6 58.0 36 9 A32693  
c 40 11.6 58.0 36 9 AX036277  
c 41 11.6 58.0 36 9 AX038887  
c 42 11.6 58.0 48 9 AR053527  
c 43 11.4 57.0 18 9 AR009347  
c 44 11.4 57.0 18 9 AR060812  
c 45 11.4 57.0 18 10 I31591

## ALIGNMENTS

RESULT 1  
AR016339/c 24 bp DNA PAT 05-DEC-1998  
LOCUS Sequence 18 from patent US 5776694.  
DEFINITION AR016339  
ACCESSION AR016339.1 GI:3972616  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 24)  
AUTHORS Sheiness,D.K., Adams,T.H., Stamm,M.R., Cangelosi,G.A.,  
Britschgi,T.B. and Dix,C.K.  
TITLE Diagnostic kits useful for selectively detecting microorganisms in  
samples  
JOURNAL Patent: US 5776694-A 18 07-JUL-1998;  
FEATURES Location/Qualifiers  
source 1..24  
BASE COUNT 7 a 7 c 6 g 4 t  
ORIGIN

Query Match 69.0%; Score 13.8; DB 9; Length 24;  
Best Local Similarity 88.2%; Pred. No. 1.8e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 ttaggatggtgccgc 19  
|| ||||| |||||  
Db 18 TTGGGATGGAGCCGTC 2

RESULT 2  
I59963/c 24 bp DNA PAT 07-OCT-1997  
LOCUS Sequence 18 from patent US 5654418.  
DEFINITION I59963  
ACCESSION I59963.1 GI:2478595  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 24)  
AUTHORS Sheiness,D.K., Cangelosi,G.A. and Britschgi,T.B.  
TITLE Nucleic acid probes useful for detecting microorganisms associated  
with vaginal infections  
JOURNAL Patent: US 5654418-A 18 05-AUG-1997;  
FEATURES Location/Qualifiers  
source 1..24  
BASE COUNT 7 a 7 c 6 g 4 t  
ORIGIN

Query Match 69.0%; Score 13.8; DB 10; Length 24;  
Best Local Similarity 88.2%; Pred. No. 1.8e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 ttaggatggtgccgc 19  
|| ||||| |||||  
Db 18 TTGGGATGGAGCCGTC 2

RESULT 3  
I86044/c 24 bp DNA PAT 10-JUN-1998  
LOCUS Sequence 18 from patent US 5700636.  
DEFINITION I86044  
ACCESSION I86044.1 GI:3205762  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 24)  
AUTHORS Sheiness,D.K., Adams,T.H., Stamm,M.R., Cangelosi,G.A.,  
Britschgi,T.B. and Dix,C.K.  
TITLE Methods for selectively detecting microorganisms associated with  
vaginal infections in complex biological samples  
JOURNAL Patent: US 5700636-A 18 23-DEC-1997;  
FEATURES Location/Qualifiers  
source 1..24  
BASE COUNT 7 a 7 c 6 g 4 t  
ORIGIN

Query Match 69.0%; Score 13.8; DB 10; Length 24;  
Best Local Similarity 88.2%; Pred. No. 1.8e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 ttaggatggtgccgc 19  
|| ||||| |||||  
Db 18 TTGGGATGGAGCCGTC 2

RESULT 4  
I24249 30 bp DNA PAT 07-OCT-1996  
LOCUS Sequence 36 from patent US 5543293.  
DEFINITION I24249  
ACCESSION I24249.1 GI:1604119  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Gold, L. and Tasset, D.  
TITLE DNA ligands of thrombin  
JOURNAL Patent: US 5343293-A 36 06-AUG-1996;  
FEATURES Location/Qualifiers  
source 1..30  
BASE COUNT 4 a 1 c 15 g 10 t  
ORIGIN

Query Match 69.0%; Score 13.8; DB 10; Length 30;  
Best Local Similarity 88.2%; Pred. No. 1.7e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggttaggagtggtgcgcg 17  
||| ||||| ||||| ||  
Db 11 GGTTAGGATGGTAGCG 27

RESULT 5  
AX001078/c 34 bp DNA PAT 10-MAR-2000  
LOCUS Sequence 3 from Patent WO9902647..-  
ACCESSION AX001078  
VERSION AX001078.1 GI:7241296  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 34)  
AUTHORS Sandig, V. and Loeser, P.  
TITLE CLONING VECTORS FOR PRODUCING ADENO VIRAL MINIMAL VIRUSES  
JOURNAL Patent: WO 9902647-A 3 21-JAN-1999;  
FEATURES HEPAVEC AG FUR GENTHERAPIE (DE); SANDIG VOLKER (DE)  
source 1..34  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 8 a 11 c 8 g 7 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 34;  
Best Local Similarity 82.4%; Pred. No. 1.2e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 tagggatggtgcgcgca 20  
||| ||||| ||||| ||  
Db 18 TAGAGTTGGTACCGTCA 2

RESULT 6  
AR087980 26 bp DNA PAT 07-SEP-2000  
LOCUS Sequence 87 from patent US 5989821.  
ACCESSION AR087980  
VERSION AR087980.1 GI:10014743  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS Goh, S.Han, Chow, A. and Hemmingsen, S.  
TITLE Universal targets for species identification  
JOURNAL Patent: US 5989821-A 87 23-NOV-1999;  
FEATURES Location/Qualifiers  
source 1..26  
/organism="unknown"  
BASE COUNT 5 a 6 c 10 g 5 t  
ORIGIN

Query Match 60.0%; Score 12; DB 9; Length 26;  
Best Local Similarity 75.0%; Pred. No. 1.6e+05;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggttaggagtggtgcgcgca 20  
||| ||||| ||||| ||  
Db 6 GGCTGGGATGGTACCACCA 25

RESULT 7  
I80198 26 bp DNA PAT 10-JUN-1998  
LOCUS Sequence 87 from patent US 5708160.  
ACCESSION I80198  
VERSION I80198.1 GI:3208488  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS Goh, S.Han, Chow, A.W. and Hemmingsen, S.  
TITLE HSP-60 genomic locus and primers for species identification  
JOURNAL Patent: US 5708160-A 87 13-JAN-1998;  
FEATURES Location/Qualifiers  
source 1..26  
/organism="unknown"  
BASE COUNT 5 a 6 c 10 g 5 t  
ORIGIN

Query Match 60.0%; Score 12; DB 10; Length 26;  
Best Local Similarity 75.0%; Pred. No. 1.6e+05;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggttaggagtggtgcgcgca 20  
||| ||||| ||||| ||  
Db 6 GGCTGGGATGGTACCACCA 25

RESULT 8  
I07710 32 bp PAT 02-DEC-1994  
LOCUS Sequence 15 from Patent EP 0362526.  
ACCESSION I07710  
VERSION I07710.1 GI:589756  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Springer, T.A. and Larson, R.  
TITLE The Alpha-subunit of the LFA-1 leukocyte adhesion receptor  
JOURNAL Patent: EP 0362526-A2 15 11-APR-1990;  
FEATURES Location/Qualifiers  
source 1..32  
/organism="unknown"  
BASE COUNT 5 a 3 c 14 g 10 t  
ORIGIN

Query Match 60.0%; Score 12; DB 10; Length 32;  
Best Local Similarity 75.0%; Pred. No. 1.6e+05;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggttaggagtggtgcgcgca 20  
||| ||||| ||||| ||  
Db 11 GGTATGGATGGTGGGCTCA 30

RESULT 9  
A92270 39 bp DNA PAT 23-JAN-2000  
LOCUS

Qy 1 ggtaggagtggtgccgtca 20  
 ||| ||| ||| ||| ||| ||| |||  
 Db 22 GGTCCGGACGGTGCCGGCA 3

RESULT 11  
 S63845/c  
 LOCUS S63845 45 bp mRNA PRI 04-MAY-2000  
 DEFINITION (D3(TM4-del))-D3 dopamine receptor {3' region, alternatively spliced} [human, peripheral blood lymphocytes, mRNA Partial Mutant, 45 nt].  
 ACCESSION S63845  
 VERSION S63845 GI:400540  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 45)  
 AUTHORS Nagai,Y., Ueno,S., Saeki,Y., Soga,F. and Yanagihara,T.  
 TITLE Expression of the D3 dopamine receptor gene and a novel variant transcript generated by alternative splicing in human peripheral blood lymphocytes  
 JOURNAL Biochem. Biophys. Res. Commun. 194 (1), 368-374 (1993)  
 MEDLINE 93326145  
 REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 135517] from the original journal article.  
 This sequence comes from Fig. 4.  
 FEATURES  
 source  
 1. 45  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="peripheral lymphocytes"  
 /tissue\_type="blood"  
 <1. 45  
 /gene="D3R"  
 /allele="D3(TM4-del)"  
 <1. 45  
 /gene="D3R"  
 /note="This sequence comes from Fig. 4; conceptual translation presented here differs from translation in publication"  
 /codon\_start=1  
 /product="D3 dopamine receptor"  
 /protein\_id="AA827543.2"  
 /db\_xref="GI:7704924"  
 /translation="SIDRGPHELLHLP"  
 11 a 17 c 9 g 8 t  
 BASE COUNT  
 ORIGIN

Query Match 60.0%; Score 12; DB 97; Length 45;  
 Best Local Similarity 75.0%; Pred. No. 1.5e+05;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggtaggagtggtgccgtca 20  
 ||||| ||||| ||||| |||||  
 Db 41 GGTGGAGATGGACGACACA 22

RESULT 12  
 E08846/c  
 LOCUS E08846 20 bp DNA PAT 29-SEP-1997  
 DEFINITION PCR primer to detect polymorphism of Histamine H1 receptor gene.  
 ACCESSION E08846  
 VERSION E08846.1 GI:2176950  
 KEYWORDS JP 1995067654-A/4.  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 unclassified.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Shimizu,S., Shimazu,M., Habano,W. and Hayashi,T..  
 TITLE HUMAN HISTAMINE H1 RECEPTOR GENE AND ITS UTILIZATION

JOURNAL Patent: JP 1995067654-A 4 14-MAR-1995;

COMMENT MITSUBISHI KAGAKU B C L:KK

OS None

OC Artificial sequences.

PN JP 1995067654-A/4

PD 14-MAR-1995

PF 03-SEP-1993 JP 1993219544

PI SHIMIZU SHOICHI, SHIMAZU MITSUNOBU, HABANO WATARU, PI

HAYASHI TOMOKO

PC C12N15/09,C12Q1/68;

CC strandedness: Single;

CC topology: Linear;

FH Key Location/Qualifiers

FT source 1..20

FT Location/Qualifiers 'Artificial sequences'.

FEATURES source

1..20 Location/Qualifiers

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 5 a 7 c 3 g 5 t

ORIGIN

Query Match 59.0%; Score 11.8; DB 10; Length 20;

Best Local Similarity 86.7%; Pred. No. 2.2e+05;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gggatgggtgcgcgtca 20

|||||

Db 19 GGGATGGTGACTTCA 5

|||||

RESULT 13

AR087898/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..26

/organism="unknown"

BASE COUNT 6 a 11 c 7 g 2 t

ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 26;

Best Local Similarity 86.7%; Pred. No. 2.1e+05;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gggatgggtgcgcgtca 20

|||||

Db 26 GTGGTGGTCCGCTCA 12

|||||

RESULT 14

AR087920/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..26

/organism="unknown"

BASE COUNT 6 a 11 c 7 g 2 t

ORIGIN

REFERENCE

1 (bases 1 to 26)

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..26

/organism="unknown"

BASE COUNT 6 a 11 c 7 g 2 t

ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 26;

Best Local Similarity 86.7%; Pred. No. 2.1e+05;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gggatgggtgcgcgtca 20

|||||

Db 26 GTGGTGGTCCGCTCA 12

|||||

RESULT 15

AR087948/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 26)

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..26

/organism="unknown"

BASE COUNT 5 a 9 c 9 g 3 t

ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 26;

Best Local Similarity 86.7%; Pred. No. 2.1e+05;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gggatgggtgcgcgtca 20

|||||

Db 26 GTGGTGGTCCGCTCA 12

|||||

Search completed: October 2, 2001, 15:56:48

Job time: 14172 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:41 ; Search time 876.95 Seconds  
(without alignments)  
14.320 Million cell updates/sec

Title: US-09-757-100B-14

Perfect score: 20

Sequence: 1 ggtaggagtggtgcgctca 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	22 AAC65546	Human focal adhesi
2	15	75.0	15	22 AAC65566	Human focal adhesi
3	14.2	71.0	34	19 AAV20799	Flea serine protea
c 4	13.8	69.0	24	13 AAQ24636	Probe Gv003 specif
c 5	13.8	69.0	24	14 AAQ52752	Sequence of probe
c 6	13.8	69.0	24	18 AAT89823	Oligonucleotide pr
c 7	13.8	69.0	24	19 AAV36792	Nucleotide sequenc
c 8	13.8	69.0	30	16 AAT00210	Thrombin DNA ligand
c 9	13.8	69.0	30	22 AAF70762	Thrombin high affi
c 10	13.4	67.0	18	20 AAZ40968	Human RhoC phospho
c 11	13.4	67.0	18	20 AAZ40969	Human RhoC phospho

c 12	13.4	67.0	18	22 AAF94689	Rho C antisense ph
c 13	13.4	67.0	18	22 AAF94690	Rho C antisense ph
c 14	13.2	66.0	33	14 AAQ36303	HMGCoA reductase t
c 15	13.2	66.0	49	20 AAZ22839	Extreme thermophil
c 16	12.8	64.0	26	14 AAQ52003	B-cell mRNA ribozy
c 17	12.6	63.0	22	19 AAV58750	PCR primer 37803S
c 18	12.6	63.0	27	17 AAT45956	24-Hydroxylase gen
c 19	12.6	63.0	33	19 AAV01423	Human T-cell recep
c 20	12.6	63.0	33	22 AAF98242	C neofomans strai
c 21	12.6	63.0	35	17 AAT45957	24-Hydroxylase gen
c 22	12.6	63.0	37	19 AAV27549	Streptococcus pneu
c 23	12.4	62.0	17	22 AAF61223	Potato dihydro-oro
c 24	12.2	61.0	20	18 AAT48710	Probe for detectin
c 25	12.2	61.0	20	20 AAV73114	Human ras oncogene
c 26	12.2	61.0	20	21 AAC60575	Human fra-1 mRNA a
c 27	12.2	61.0	30	21 AAZ93264	PCR primer h/mPCR:
c 28	12.2	61.0	34	20 AAX02783	Vector pmwX-BG oli
c 29	12	60.0	15	22 AAF51105	IGF-I oligonucleot
c 30	12	60.0	15	22 AAF51106	IGF-I oligonucleot
c 31	12	60.0	15	22 AAF51107	IGF-I oligonucleot
c 32	12	60.0	15	22 AAF51108	IGF-I oligonucleot
c 33	12	60.0	26	17 AAT69048	Primer for univers
c 34	12	60.0	36	12 AAQ10384	Untranslated regio
c 35	12	60.0	37	20 AAZ08497	Competitor oligonu
c 36	12	60.0	37	20 AAZ08500	G-free binding ele
c 37	12	60.0	39	21 AAZ37033	Calcineurin bindin
c 38	11.8	59.0	20	16 AAQ87917	Human histamine H1
c 39	11.8	59.0	23	21 AAA71892	B. cereus Yisi hom
c 40	11.8	59.0	26	17 AAT69038	Primer for univers
c 41	11.8	59.0	26	17 AAT69040	Primer for univers
c 42	11.8	59.0	26	17 AAT69046	Primer for univers
c 43	11.8	59.0	26	17 AAT68966	Primer for univers
c 44	11.8	59.0	26	17 AAT69015	Primer for univers
c 45	11.8	59.0	26	17 AAT68984	Primer for univers

#### ALIGNMENTS

#### RESULT 1

AC65546  
ID AAC65546 standard; DNA; 20 BP.

XX AAC65546;

XX 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #12.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.

XX Homo sapiens.

XX US6133031-A.

PD 17-OCT-2000.

XX 19-AUG-1999; 99US-0377310.

XX 19-AUG-1999; 99US-0377310.

XX (ISIS-) ISIS PHARM INC.

XX Monia BP, Gaarde WA;

XX WPI; 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -

PS Claim 15; Column 23; 30pp; English.

CC The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, embryonic  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

SQ Sequence 20 BP; 3 A; 3 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.35; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtagggatggtgcccgtca 20  
|||||

Db 1 ggtagggatggtgcccgtca 20  
|||||

RESULT 2

AA655566  
ID AAC65566 standard; DNA; 15 BP.

AC AAC65566;

XX 12-FEB-2001 (first entry)

DE Human focal adhesion kinase antisense sequence #32.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.

OS Homo sapiens.

XX US6133031-A.

XX 17-OCT-2000.

PF 19-AUG-1999; 99US-0377310.

PR 19-AUG-1999; 99US-0377310.

XX (ISIS-) ISIS PHARM INC.

XX Monia BP, Gaarde WA;

XX WPI: 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -

PS Example 2; Column 25; 30pp; English.

CC The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, embryonic  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

SQ Sequence 15 BP; 2 A; 2 C; 7 G; 4 T; 0 other;

Query Match 75.0%; Score 15; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ttagggatggtgccc 17  
|||||

Db 1 ttagggatggtgccc 15  
|||||

RESULT 3

AAV20799  
ID AAV20799 standard; DNA; 34 BP.

XX AAV20799;

XX 02-JUL-1998 (first entry)

DE Flea serine protease PCR primer SEQ ID NO:58.

XX Flea: Siphonaptera; protease; vaccine; prophylaxis; infestation;  
KW Immunoglobulin protease; larvae; host animal; PCR primer; ss.

XX Synthetic.

OS Siphonaptera.

XX WO9740058-A1.

XX 30-OCT-1997.

XX 24-APR-1997; 97WO-US06121.

XX 04-APR-1997; 97US-0042945.

XX 24-APR-1996; 96US-0639075.

XX 15-NOV-1996; 96US-0749699.

XX (HESK-) HESKA CORP.

XX Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE;

PI Silver G, Stiegler GL;

XX WPI: 1998-076762/07.

XX New flea protease genes and proteins - used in vaccine compositions  
PT for the prophylaxis and treatment of flea infestation, especially in  
PT cats or dogs

XX Example 23; Page 217; 318pp; English.

CC The present sequence represents a PCR primer for flea serine protease.  
CC The protease, its mimetopes, antibodies (Ab) and inhibitors of the  
CC protein, as well as the DNA encoding the protein, may all be used in  
CC therapeutic compositions to reduce flea protease activity (especially in  
CC immunoglobulin protease) and so reduce flea infestation, especially in  
CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult  
CC fleas which comprises anti-protease antibodies produced by a host animal  
CC in response to administration of the protein. Therapeutic compositions  
CC may further comprise a compound that reduces haematophagous ectoparasite  
CC burden by a method other than by reducing flea immunoglobulin protease  
CC activity. The novel flea DNA encoding the protein can also be used to  
CC produce recombinant protein, and fragments of it are used as probes and  
CC primers for identification and isolation of related sequences, also as  
CC antisense, triplex-forming agents and ribozymes for inhibition of the  
CC synthesis of the protein. Ab are also useful for screening expression  
CC libraries, to purify the protein and to target cytotoxins to fleas.

XX Sequence 34 BP; 7 A; 11 C; 8 G; 8 T; 0 other;

Query Match 71.0%; Score 14.2; DB 19; Length 34;

Best Local Similarity 84.2%; Pred. No. 2.9e+02;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttaggatggtgcccgtca 20  
|||||

Db 13 gttaggatggtgcccgtca 31  
|||||



RESULT 4  
AAQ24636/c  
ID AAQ24636 standard; DNA; 24 BP.

XX AC AAQ24636;  
XX DT 05-NOV-1992 (first entry)

XX DE Probe Gv003 specific for Gardnerella vaginalis.

XX KW Dipstick; capture probe; 16S ribosomal RNA; vaginal infection; ss.

XX OS Synthetic.

XX PN W09207096-A.

XX PD 30-APR-1992.

XX PF 21-OCT-1991; 91WO-US07763.

XX PR 19-OCT-1990; 90US-0600334.

XX PA (MICR-) MICROPROBE CORP.

XX PI Adams TH, Sheiness DK;

XX DR WPI; 1992-167174/20.

XX DT Detecting bacterial vaginitis infections - by contacting formed

XX PT lysate with hybridisation soln. and dipstick comprising

XX PT non-porous solid support and determining signal associated with

XX PT beads

XX PS Disclosure; Page 39; 62pp; English.

XX CC This oligonucleotide is one example of a probe suitable for

XX CC diagnosing the presence of G.vaginalis. The preferred target for

XX CC such probes is 16S rRNA, present in several thousand copies per cell.

XX CC Oligonucleotides complementary to sequences in the G.vaginalis

XX CC genome or in G.vaginalis plasmids may also be used.

XX CC See also AAQ24637-Q24647.

XX SQ Sequence 24 BP; 7 A; 7 C; 6 G; 4 T; 0 other;

Query Match 69.0%; Score 13.8; DB 13; Length 24;

Best Local Similarity 88.2%; Pred. No. 4.5e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ttagggtggtggtcgcgc 19

DB 18 TTTGGGATGGAGCCGTC 2

RESULT 5

AAQ52752/c

ID AAQ52752 standard; DNA; 24 BP.

XX AC AAQ52752;

XX DT 24-JUN-1994 (first entry)

XX DE Sequence of probe Gv003 for the detection of Gardnerella vaginalis.

XX KW Probe; vaginal fluid microorganism; oligonucleotide; gram positive;

XX KW ss.

XX OS Synthetic.

XX PN W09324659-A.

XX PD 09-DEC-1993.

PF 28-MAY-1993; 93WO-US05085.

XX PR 29-MAY-1992; 92US-0896094.

XX PA (MICR-) MICROPROBE CORP.

XX PI Adams TH, Britschgi TB, Cangelosi GA, Sheiness DK;

XX PI Stamm MR;

XX DR WPI; 1993-405847/50.

XX DE Detection of microorganisms, partic. in vaginal fluid samples -

XX PT by lysis of cells to release nucleic acid and hybridisation

XX PT assays using probes

XX PS Claim 9; Page 42; 78pp; English.

XX CC The inventors claim a compsn. of at least one oligo probe for the

XX CC detection of at least one microorganism found in vaginal samples.

XX CC The sequence below is one such probe. The methods and compsns. can

XX CC be used for the detection of microorganisms, partic. in vaginal

XX CC fluid samples.

XX SQ Sequence 24 BP; 7 A; 7 C; 6 G; 4 T; 0 other;

Query Match 69.0%; Score 13.8; DB 14; Length 24;

Best Local Similarity 88.2%; Pred. No. 4.5e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ttagggtggtggtcgcgc 19

DB 18 TTTGGGATGGAGCCGTC 2

RESULT 6

AAT89823/c

ID AAT89823 standard; DNA; 24 BP.

XX AC AAT89823;

XX DT 27-FEB-1998 (first entry)

XX DE Oligonucleotide probe Gv003 specific for Gardnerella vaginalis.

XX KW Probe; hybridise; ribosomal RNA; vaginitis; infection; diagnostic;

XX KW ribosomal RNA; ss.

XX OS Synthetic.

XX OS Gardnerella vaginalis.

XX PN US5654418-A.

XX PD 05-AUG-1997.

XX PF 02-JUN-1995; 95US-0460344.

XX PR 08-OCT-1993; 93US-0133598.

XX PR 19-OCT-1990; 90US-0600334.

XX PR 29-MAY-1992; 92US-0896094.

XX PA (BECT ) BECTON DICKINSON CO.

XX PI Britschgi TB, Cangelosi GA, Sheiness DK;

XX DR WPI; 1997-401906/37.

XX PT Oligo:nucleotide probes for detecting microorganisms in vaginal

XX PT samples - allows selective detection of vaginitis-causing

XX PT microorganisms

XX PS Claim 1; Column 32; 40pp; English.



Db 11 ggtagggatggtgacg 27  
|||||

## RESULT 9

AAF70762  
ID AAF70762 standard; DNA; 30 BP.

XX  
AC AAF70762;

DT 20-APR-2001 (first entry)

DE Thrombin high affinity ligand #9.

XX Ligand: basic fibroblast growth factor; bFGF; gene therapy; vascular;  
KW atherosclerosis; angioplasty; stability; ss.

XX Unidentified.

PN US6177557-B1.

PD 23-JAN-2001.

PF 05-AUG-1996; 96US-0687421.

PR 11-JUN-1990; 90US-0536428.

PR 10-JUN-1991; 91US-0714131.

PR 06-NOV-1992; 92US-0973333.

PR 10-FEB-1994; 94US-0195005.

PR 28-MAR-1994; 94US-0219012.

XX (NEXS-) NEXSTAR PHARM INC.

XX Janjic N, Gold L, Tasset D;

XX WPI; 2001-158583/16.

PT Novel nucleic acid ligands to basic fibroblast growth factor that are  
PT useful as inhibitors of basic fibroblast growth factors and 2'-amino  
PT modified RNA ligands, exhibit increased in vivo stability -

XX Example 19; Column 57-58; 153pp; English.

XX The present invention relates to a purified and isolated non-naturally  
CC occurring DNA ligands to basic fibroblast growth factor (bFGF).  
CC The ligands are useful as part of gene therapy treatments and  
CC for diagnosing pathogenesis of vascular diseases including  
CC initiation and progression of atherosclerosis, acute coronary  
CC syndromes, vein graft disease and restenosis following coronary  
CC angioplasty. The ligands have improved stability in vivo.

XX Sequence 30 BP; 4 A; 1 C; 15 G; 10 T; 0 other;

Query Match 69.0%; Score 13.8; DB 22; Length 30;

Best Local Similarity 88.2%; Pred. No. 4.6e-02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggtagggatggtgacg 17  
|||||

Db 11 ggtagggatggtgacg 27  
|||||

## RESULT 10

AAZ40968/c  
ID AAZ40968 standard; DNA; 18 BP.

XX  
AC AAZ40968;

DT 26-JAN-2000 (first entry)

XX Human RhoC phosphorothioate antisense oligonucleotide SEQ ID NO:120.

XX

KW Identification; genetic target; gene modulation; human; probe;  
KW antisense oligonucleotide; phosphorothioate; PCR primer;  
KW nucleotide sequence-based technology; antisense drug discovery;  
KW target validation; ss.

XX Synthetic.

OS Homo sapiens.

XX WO9953101-A1.

XX PD 21-OCT-1999.

XX PF 13-APR-1999; 99WO-US08268.

XX PR 13-APR-1998; 98US-0081483.

XX PR 28-APR-1998; 98US-0067638.

XX (ISIS-) ISIS PHARM INC.

XX Cowsert LM, Baker BF, McNeil J, Freier SM, Sasmor HM, Brooks DG;  
PI Ohasi C, Wyatt JR, Borchers AH, Vickers TA;

XX WPI; 1999-620446/53.

XX Identifying compounds which modulate expression of nucleic acids, used  
PT to provide compounds having defined physical, chemical or bioactive  
PT properties, e.g. antisense activity -

XX Example 18; Page 96; 264pp; English.

XX A method has been developed of defining a set of compounds that modulate  
CC the expression of a target nucleic acid (tNA) sequence via binding of  
CC the compounds with the tNA sequence. The method comprises generating a  
CC library of virtual compounds in silico according to defined criteria,  
CC and evaluating in silico the binding of the virtual compounds with the  
CC tNA according to defined criteria. Also described are: (1) a method of  
CC defining a set of oligonucleotides (ONs) that modulate the expression of  
CC a tNA sequence via binding of the ONs with the tNA sequence comprising  
CC generating a library of virtual compounds in silico according to defined  
CC criteria, and evaluating in silico the binding of the virtual ONs with  
CC the tNA according to defined criteria; and (2) a method of defining a  
CC set of compounds that modulate the expression of a tNA sequence via  
CC binding of the compounds with the tNA. The methods can be used for the  
CC generation and identification of synthetic compounds having defined  
CC physical, chemical or bioactive properties. Information gathered from  
CC assays of such compounds is used to identify nucleic acid sequences that  
CC are tractable to a variety of nucleotide sequence-based technologies,  
CC e.g. antisense drug discovery and target validation. AAZ40852 to  
CC AAZ41220, and AAY52701 to AAY52706, represent sequences used in the  
CC exemplification of the present invention.

XX Sequence 18 BP; 5 A; 8 C; 3 G; 2 T; 0 other;

Query Match 67.0%; Score 13.4; DB 20; Length 18;

Best Local Similarity 93.3%; Pred. No. 7e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gttagggatggtgcc 16  
|||

Db 15 GTTGGGGATGGTGC 1  
|||

## RESULT 11

AAZ40969/c  
ID AAZ40969 standard; DNA; 18 BP.

XX  
AC AAZ40969;

DT 26-JAN-2000 (first entry)

XX Human RhoC phosphorothioate antisense oligonucleotide SEQ ID NO:121.

XX

KW Identification; genetic target; gene modulation; human; probe;  
 KW antisense oligonucleotide; phosphorothioate; PCR primer;  
 KW nucleotide sequence-based technology; antisense drug discovery;  
 KW target validation; ss.

XX Synthetic.  
 OS Homo sapiens.

XX WO9953101-A1.

XX 21-OCT-1999.

XX 13-APR-1999; 99WO-US08268.

XX 13-APR-1998; 98US-0081483.

XX 28-APR-1998; 98US-0067638.

XX (ISIS-) ISIS PHARM INC.

PI Cowsett LM, Baker BF, McNeil J, Freier SM, Sasmor HM, Brooks DG;

PI Ohasi C, Wyatt JR, Borchers AH, Vickers TA;

XX WPI; 1999-620446/53.

XX Identifying compounds which modulate expression of nucleic acids, used  
 PT to provide compounds having defined physical, chemical or bioactive  
 PT properties, e.g. antisense activity

XX Example 18; Page 96; 264pp; English.

XX A method has been developed of defining a set of compounds that modulate  
 CC the expression of a target nucleic acid (tNA) sequence via binding of  
 CC the compounds with the tNA sequence. The method comprises generating a  
 CC library of virtual compounds in silico according to defined criteria,  
 CC and evaluating in silico the binding of the virtual compounds with the  
 CC tNA according to defined criteria. Also described are: (1) a method of  
 CC defining a set of oligonucleotides (ONS) that modulate the expression of  
 CC a tNA sequence via binding of the ONS with the tNA sequence comprising  
 CC generating a library of virtual compounds in silico according to defined  
 CC criteria, and evaluating in silico the binding of the virtual ONS with  
 CC the tNA according to defined criteria; and (2) a method of defining a  
 CC set of compounds that modulate the expression of a tNA sequence via  
 CC binding of the compounds with the tNA. The methods can be used for the  
 CC generation and identification of synthetic compounds having defined  
 CC physical, chemical or bioactive properties. Information gathered from  
 CC assays of such compounds is used to identify nucleic acid sequences that  
 CC are tractable to a variety of nucleotide sequence-based technologies,  
 CC e.g. antisense drug discovery and target validation. AAZ40852 to  
 CC AAZ41220, and AAZ52701 to AAZ52706, represent sequences used in the  
 CC exemplification of the present invention.

XX Sequence 18 BP; 6 A; 8 C; 3 G; 1 T; 0 other;

Query Match 67.0%; Score 13.4; DB 20; Length 18;  
 Best Local Similarity 93.3%; Pred. No. 7e+02;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gttaggagtggtgcc 16  
 ||| |||||  
 DB 16 GTTGGGATGGTGCC 2

RESULT 12

AAAF94689/C  
 ID AAF94689 standard; DNA; 18 BP.

XX AAF94689;

XX 23-MAY-2001 (first entry)

XX Rho C antisense phosphorothioate oligonucleotide SEQ ID 113.

KW Rho; GTP binding protein; phosphorothioate antisense oligonucleotide;  
 KW RhoA; RhoB; RhoC; RhoG; Rac 1; cdc42; hyperproliferative condition;  
 KW cancer; wound healing; clotting; ischaemia; reperfusion; reoxygenation;  
 KW ss.

XX Homo sapiens.

XX WO200115739-A1.

XX 08-MAR-2001.

XX 18-AUG-2000; 2000WO-US22808.

XX 31-AUG-1999; 99US-0387341.

XX (ISIS-) ISIS PHARM INC.

XX Roberts ML, Cowsett LM;

XX WPI; 2001-191677/19.

XX An antisense compound targeted to a nucleic acid molecule encoding a  
 PT member of the human Rho family of small GTP binding proteins useful for  
 PT treating e.g. cancer and ischaemia

XX Example 16; Page 73; 156pp; English.

XX This invention relates to an antisense compound targeted to a nucleic  
 CC acid molecule encoding a member of the human Rho family of small GTP  
 CC binding proteins, where the antisense compound inhibits the expression of  
 CC the member of the human Rho family. The invention includes antisense  
 CC oligonucleotides AAF94580 - AAF94637 which target a RhoA nucleotide  
 CC sequence, AAF94645 - AAF94684 which target a RhoB nucleotide sequence,  
 CC AAF94686 - AAF94725 which target a RhoC nucleotide sequence, AAF94727 -  
 CC AAF94766 which target RhoG nucleotide sequence, AAF94769 - AAF94790 which  
 CC target a Rac 1 nucleotide sequence and AAF94795 - AAF94809 which target  
 CC cdc42 nucleotide sequence. The antisense compound is useful for treating  
 CC hyperproliferative conditions, especially cancer, abnormal wound healing  
 CC or clotting conditions and ischaemia/reperfusion or reoxygenation injury.  
 CC The compound may also be used to diagnose the above conditions.

XX Sequence 18 BP; 5 A; 8 C; 3 G; 2 T; 0 other;

Query Match 67.0%; Score 13.4; DB 22; Length 18;  
 Best Local Similarity 93.3%; Pred. No. 7e+02;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gttaggagtggtgcc 16  
 ||| |||||  
 DB 15 GTTGGGATGGTGCC 1

RESULT 13

AAAF94690/C

ID AAF94690 standard; DNA; 18 BP.

XX AAF94690;

XX 23-MAY-2001 (first entry)

XX Rho C antisense phosphorothioate oligonucleotide SEQ ID 114.

KW Rho; GTP binding protein; phosphorothioate antisense oligonucleotide;  
 KW RhoA; RhoB; RhoC; RhoG; Rac 1; cdc42; hyperproliferative condition;  
 KW cancer; wound healing; clotting; ischaemia; reperfusion; reoxygenation;  
 KW ss.

XX Homo sapiens.

XX WO200115739-A1.

XX 08-MAR-2001.

XX 18-AUG-2000; 2000WO-US22808.  
 XX 31-AUG-1999; 99US-0387341.  
 XX (ISIS-) ISIS PHARM INC.  
 XX Roberts ML, Cowsert LM;  
 XX WPI; 2001-191677/19.  
 XX An antisense compound targeted to a nucleic acid molecule encoding a  
 PT member of the human Rho family of small GTP binding proteins useful for  
 PT treating e.g. cancer and ischaemia -  
 XX Example 16; Page 73; 156pp; English.  
 XX This invention relates to an antisense compound targeted to a nucleic  
 CC acid molecule encoding a member of the human Rho family of small GTP  
 CC binding proteins, where the antisense compound inhibits the expression of  
 CC the member of the human Rho family. The invention includes antisense  
 CC oligonucleotides AAF94580 - AAF94637 which target a RhoA nucleotide  
 CC sequence, AAF94645 - AAF94684 which target a RhoB nucleotide sequence,  
 CC AAF94686 - AAF94725 which target a RhoC nucleotide sequence, AAF94727 -  
 CC AAF94766 which target RhoG nucleotide sequence, AAF94769 - AAF94790 which  
 CC target a Rac 1 nucleotide sequence and AAF94795 - AAF94809 which target  
 CC cdc42 nucleotide sequence. The antisense compound is useful for treating  
 CC hyperproliferative conditions, especially cancer, abnormal wound healing  
 CC or clotting conditions and ischaemia/reperfusion or reoxygenation injury.  
 CC The compound may also be used to diagnose the above conditions.  
 XX Sequence 18 BP; 6 A; 8 C; 3 G; 1 T; 0 other;

Query Match 67.0%; Score 13.4; DB 22; Length 18;  
 Best Local Similarity 93.3%; Pred. No. 7e+02;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gttagggtggtgcc 16  
 ||| |||||  
 DB 16 GTTGGGGTGGTGC 2

RESULT 14  
 AAQ36303  
 ID AAQ36303 standard; DNA; 33 BP.  
 XX  
 AC AAQ36303;  
 XX  
 DT 07-JUN-1993 (first entry)  
 XX  
 DE HMGCoA reductase target sequence.  
 XX  
 KW Cholesterol; repressor protein; transcription; agonist; cellular;  
 KW triplex; target; duplex; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN US5176996-A.  
 XX  
 PD 05-JAN-1993.  
 XX  
 PF 20-DEC-1988; 88US-0287359.  
 XX  
 PR 20-DEC-1988; 88US-0287359.  
 PR 22-DEC-1989; 89US-0453532.  
 XX  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 XX  
 PI Hogan ME, Kessler DJ;  
 XX  
 DR WPI; 1993-035718/04.  
 XX

PT Synthetic oligo-nucleotide(s), prodn. useful e.g. for HIV-1  
 PT inhibition - which bind to target sequence in duplex DNA forming  
 PT colinear triplex by binding to major groove  
 XX  
 PS Example 9; Column 28; 29pp; English.  
 XX  
 CC HMGCoA reductase is the enzyme which defines the rate limiting step  
 CC in cholesterol biosynthesis. Expression of the gene may be  
 CC prevented by the formation of a triplex between the duplex target DNA  
 CC sequence and an anti parallel or parallel synthetic oligonucleotide.  
 CC A suitable target sequence is the binding site for a repressor  
 CC protein which appears to mediate end-prod. inhibition of transcription  
 CC by cholesterol, from base -167 to -135 of HMGCoA reductase gene.  
 CC Oligonucleotides targetted against this sequence will repress  
 CC HMGCoA expression, and act as agonists of the cellular repressor.  
 CC See also AAQ36219-362.  
 XX  
 SQ Sequence 33 BP; 3 A; 10 C; 13 G; 7 T; 0 other;

Query Match 66.0%; Score 13.2; DB 14; Length 33;  
 Best Local Similarity 83.3%; Pred. No. 9.3e+02;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggttagggatggtgcgt 18  
 ||| |||||  
 DB 1 ggtgagagatggtgcgt 18

RESULT 15  
 AAZ22839/c  
 ID AAZ22839 standard; DNA; 49 BP.  
 XX  
 AC AAZ22839;  
 XX

DT 06-DEC-1999 (first entry)  
 XX

DE Extreme thermophilic mismatch cleavage enzyme substrate oligo #4.

KW Thermophilic; mismatch; cleavage; PCR; amplification; Taq polymerase;  
 KW mutation; detection; pathogen; cancer; probe; ss.  
 XX

OS Synthetic.

PN WO9942595-A1.  
 XX

PD 26-AUG-1999.  
 XX

PF 19-FEB-1999; 99WO-US03274.  
 XX

PR 19-FEB-1998; 98US-0075194.  
 XX

PA (TREV-) TREVIGEN INC.  
 XX

PI Chirikjian JG, Bazar LS, George AL;  
 XX

DR WPI; 1999-571571/48.  
 XX

PT New isolated extreme thermophilic enzyme, which cleaves at a mismatch  
 PT formed between two polynucleotides in a duplex -  
 XX

PS Example 4; Fig 2; 41pp; English.

XX This sequence represents oligonucleotide #4 which was annealed to  
 CC oligonucleotides #1, #2 and #3 (AAZ22836-22838) to generate a four-way  
 CC junction used to assay for any resolvase activity in an extreme  
 CC thermophilic enzyme composition from Thermotoga maritima. Extreme  
 CC thermophilic enzymes, such as TM-Endo V (from Thermotoga maritima)  
 CC cleave at a mismatch formed between two polynucleotides in a duplex. It  
 CC does not exhibit resolvase activity and does not require a GATC sequence  
 CC to effect cleavage. TM-Endo V cleaves C/C, G/G, T/C, A/C, A/A and T/T  
 CC mismatch. The enzyme does not cleave a perfectly complementary  
 CC oligonucleotide pair (no mismatch), a T/G mismatch or a 3 base bubble. It



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:00 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-14

Perfect score: 20

Sequence: 1 ggttagggatggtgcgcgtca 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1028115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0429 row: K column: 17  
Seq primer: CACACAGGAAACAGTATGACC  
Class: plasmid ends  
High quality sequence stop: 27.  
Location/Qualifiers  
1. .27  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUC1M0429K17"  
/clone\_lib="Mouse 10kb plasmid UUC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 9 c 6 g 7 t  
ORIGIN  
Query Match 71.0%; Score 14.2; DB 246; Length 27;  
Best Local Similarity 84.2%; Pred. No. 6.7e+03;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ggttaggagtggtgcgcgc 19  
|||  
Db 26 GGCTAGGATGGCCCATC 8

RESULT 2  
AZ592004 29 bp DNA GSS 13-DEC-2000  
LOCUS IM0402H06R Mouse 10kb plasmid UUC1M library Mus musculus genomic  
DEFINITION clone UUC1M0402H06 R, DNA sequence.  
ACCESSION AZ592004  
VERSION AZ592004.1 GI:11714194  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 29)  
REFERENCE  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	14.2	71.0	27	246	AZ607381 IM0429K17
c 2	13	65.0	29	246	AZ592004 IM0420H06
c 3	12.8	64.0	45	1	AA009110 mg92q09.r
c 4	12.2	61.0	27	246	AZ607381 IM0429K17
c 5	12.2	61.0	49	8	AA501145 vh73d08.r
c 6	12	60.0	23	243	AZ434508 IM0220K22
c 7	12	60.0	47	249	AZ769994 IM0571A14
c 8	11.8	59.0	37	249	AZ760007 IM0553C23
c 9	11.8	59.0	37	249	AZ760007 IM0553C23
c 10	11.6	58.0	20	242	AZ345513 IM0080J04
c 11	11.6	58.0	40	17	AI188838 qd21f12.x
c 12	11.6	58.0	41	158	H84363 yv85c09.sl
c 13	11.4	57.0	21	243	AZ399400 IM0165C13
c 14	11.4	57.0	21	249	AZ786429 IM0031L23
c 15	11.4	57.0	38	250	AZ822907 IM0096M03
c 16	11.2	56.0	25	246	AZ611815 IM0438F21
c 17	11.2	56.0	34	24	AI721161 as73h12.x
c 18	11.2	56.0	38	159	H99519 yx29b12.sl
c 19	11.2	56.0	43	244	AZ481738 IM0306N10
c 20	11.2	56.0	48	4	AA229879 nc36f11.r
c 21	11.2	56.0	50	107	AU103658 AU103658
c 22	11.2	56.0	50	107	AU105458 AU105458
c 23	11.2	56.0	50	107	AU105460 AU105460
c 24	11.2	56.0	50	107	AU105464 AU105464
c 25	11.2	56.0	50	107	AU105467 AU105467
c 26	11.2	56.0	50	107	AU105468 AU105468
c 27	11.2	56.0	50	107	AU105469 AU105469
c 28	11	55.0	45	15	AI025423 ow20h08.s
c 29	11	55.0	44	111	AW063399 DF1042.RR
c 30	11	55.0	44	155	C00309 HUMGS000601
c 31	11	55.0	44	246	AZ579484 IM0367B06
c 32	11	55.0	46	246	AZ624124 IM0462J03
c 33	11	55.0	48	246	AZ576103 AST-T29E0
c 34	11	55.0	49	3	AA213878 zr91b05.r
c 35	11	55.0	49	19	AI361993 qv49b08.x
c 36	11	55.0	49	23	AI654650 tq93a05.x
c 37	11	55.0	50	107	AU103102 AU103102
c 38	11	55.0	50	107	AU106713 AU106713
c 39	10.8	54.0	21	243	AZ399828 IM0165A23
c 40	10.8	54.0	22	247	AZ647408 IM0513H19
c 41	10.8	54.0	30	250	AZ833673 IM0115I23
c 42	10.8	54.0	34	10	AA654488 ag53f04.s
c 43	10.8	54.0	38	249	AZ806456 IM0068E05
c 44	10.8	54.0	40	258	TA130E05P
c 45	10.8	54.0	42	188	R97402 yq53b07.sl

ALIGNMENTS

RESULT 1  
AZ607381/c 27 bp DNA GSS 13-DEC-2000  
LOCUS IM0429K17R Mouse 10kb plasmid UUC1M library Mus musculus genomic  
DEFINITION clone UUC1M0429K17 R, DNA sequence.  
ACCESSION AZ607381  
VERSION AZ607381.1 GI:11729571  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 27)  
REFERENCE  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.

TITLE and Wright, D., Weiss, R.  
 JOURNAL Mouse whole genome scaffolding with paired end reads from 10kb  
 COMMENT plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: dunn@genetics.utah.edu  
 Insert length: 10000 Std Error: 0.00  
 Plate: 0402 row: H column: 06  
 Seq primer: CACACAGAAACGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 29.  
 Location/Qualifiers

## FEATURES

source

1. .29  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0402H06"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114/gb/AP129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 7 c 10 g 7 t  
 ORIGIN

Query Match 65.0%; Score 13; DB 246; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+04;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gatggtccctca 20  
 |||||  
 Db 2 GATGGTCCGCTCA 14

RESULT 3  
 AA009110/c  
 LOCUS 45 bp mRNA EST 26-JUL-1996  
 DEFINITION mg92g09.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA  
 clone IMAGE:440512 5' similar to gb:X05021 Murine mRNA with  
 homology to yeast L29 ribosomal protein (MOUSE);, mRNA sequence.  
 AA009110  
 accession AA009110.1 GI:1465176  
 version AA009110  
 keywords EST.  
 source house mouse.  
 organism Mus musculus  
 reference Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 45)  
 authors Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

TITLE  
JOURNAL  
COMMENT

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lannon, G., Soares, B., Wilson, R. and Waterston, R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:265848

Trace considered overall poor quality

Seq primer: Etpriimer

High quality sequence stop: 1.

## FEATURES

source

Location/Qualifiers  
 1. .45  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:440512"  
 /clone\_lib="Soares mouse embryo NbME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TCTTACCAATCGAAGTGGAGCGCGGAAATTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 15 a 14 c 8 g 8 t  
 ORIGIN

Query Match 64.0%; Score 12.8; DB 1; Length 45;

Best Local Similarity 87.5%; Pred. No. 3.4e+04;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggttaggagtggtgcc 16  
 |||||  
 Db 43 GGTTTGGATGGTGC 28

## RESULT

4

AZ607381

LOCUS

DEFINITION AZ607381 27 bp DNA GSS 13-DEC-2000  
 clone UUGC1M0429K17 R, DNA sequence.

ACCESSION

AZ607381

VERSION

GSS.

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

1 (bases 1 to 27)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL  
COMMENT

plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0429 row: K column: 17  
 Seq primer: CACACGAGAAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 27.

## FEATURES

Location/Qualifiers  
 1..27  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0429K17"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gii14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

5 a 9 c 6 g 7 t

Query Match 61.0%; Score 12.2; DB 246; Length 27;  
 Best Local Similarity 82.4%; Pred. No. 6.4e+04;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttagggatgggtgcgc 19  
 ||| ||||| ||| ||  
 Db 3 TTATGGATGGGCCATC 19

## RESULT 5

AA501145 49 bp mRNA EST 01-JUL-1997  
 LOCUS vH3d08.r1 Knowles Solter mouse inner cell mass Mus musculus cDNA  
 DEFINITION clone IMAGE:892623 5' similar to WP:B0491.1 CE02103 ;, mRNA sequence.

ACCESSION AA501145.1 GI:2236112  
 VERSION AA501145  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 49)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

TITLE  
JOURNAL  
COMMENT

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:520583

Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 High quality sequence stop: 1.

FEATURES  
source

Location/Qualifiers  
 1..49  
 /organism="Mus musculus"  
 /strain="B6D2 F1/J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:892623"  
 /clone\_lib="Knowles Solter mouse inner cell mass"  
 /dev\_stage="embryo (post-implantation)"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript SK+; Site1: XbaI; Site2: XhoI  
 : Cloned unidirectionally from mRNA prepared from primitive streak embryonic tissue. Primer: Oligo dT.  
 cDNAs were cloned into the XbaI/XhoI sites of pBluescript SK+ (Stratagene) using commercial linkers (NEB). Average insert size: 0.5 kb."

BASE COUNT 14 a 9 c 13 g 13 t  
 ORIGIN

Query Match 61.0%; Score 12.2; DB 8; Length 49;  
 Best Local Similarity 82.4%; Pred. No. 6.6e+04;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggtagggatgggtgcgc 17  
 ||||| ||||| ||| ||  
 Db 21 GGTAGAGATGATGCGC 37

RESULT 6  
A434508/c

AZ434508 23 bp DNA GSS 03-OCT-2000  
 LOCUS 1M0220K22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC1M0220K22 R, DNA sequence.

ACCESSION AZ434508.1 GI:10559521  
 VERSION AZ434508  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 23)  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00  
 Plate: 0220 row: K column: 22  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 23.  
 Location/Qualifiers

#### FEATURES

source  
 1. .23  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0220R22"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 9 a 8 c 4 g 2 t  
 ORIGIN

Query Match 60.0%; Score 12; DB 243; Length 23;  
 Best Local Similarity 75.0%; Pred. No. 8e+04;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggtaggagtggtgcgcgtca 20  
 ||||| | ||||| |||  
 Db 20 GGTTGCTGTGGTGCCTCA 1

#### RESULT 7

AZ769994  
 LOCUS AZ769994 47 bp DNA GSS 16-FEB-2001  
 DEFINITION IM0571A14F Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0571A14 F, DNA sequence.  
 ACCESSION AZ769994  
 VERSION AZ769994.1 GI:12890717  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 47)  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0571 row: A column: 14  
 Seq primer: CGTTGTAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 47.  
 Location/Qualifiers

#### FEATURES

source  
 1. .47  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0571A14"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 8 c 25 g 9 t  
 ORIGIN

Query Match 60.0%; Score 12; DB 249; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 8.3e+04;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 agggatgggtgcc 16  
 ||||| |||||  
 Db 21 AGGATGGTGCC 32

#### RESULT 8

AZ760007  
 LOCUS AZ760007 37 bp DNA GSS 16-FEB-2001  
 DEFINITION IM0553C23F Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0553C23 F, DNA sequence.  
 ACCESSION AZ760007  
 VERSION AZ760007.1 GI:12867379  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 37)  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606

Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0553 row: C column: 23  
 Seq primer: CGTGTAAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 37.  
 Location/Qualifiers  
 1. 37  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0553C23"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g1473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 11 a 6 c 7 g 13 t  
 ORIGIN

Query Match 59.0%; Score 11.8; DB 249; Length 37;  
 Best Local Similarity 86.7%; Pred. No. 1e+05;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggtagggatggtgc 15  
 Db 5 GGTGTGGATGTC 19

RESULT 9  
 AUI03569/c  
 LOCUS  
 DEFINITION AUI03569 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 CAS00968, mRNA sequence.  
 ACCESSION AUI03569  
 VERSION AUI03569.1 GI:13553090  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 50)  
 Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isoqai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.  
 Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
 Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: yusuki@iems.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano

FEATURES  
 source  
 1. 50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CAS00968"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
 4 a 14 c 22 g 10 t  
 BASE COUNT 4 a 14 c 22 g 10 t  
 ORIGIN

Query Match 59.0%; Score 11.8; DB 107; Length 50;  
 Best Local Similarity 86.7%; Pred. No. 1e+05;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gggatgggtgcctca 20  
 Db 39 GGGATGCTGCCGACA 25

RESULT 10  
 A2345513/c  
 LOCUS  
 DEFINITION A2345513 20 bp DNA GSS 29-SEP-2000  
 clone UUGC1M0080J04 F, DNA sequence.  
 ACCESSION A2345513  
 VERSION A2345513.1 GI:10424750  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 20)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0080 row: J column: 04  
 Seq primer: CGTGTAAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 20.  
 Location/Qualifiers  
 1. 20  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0080J04"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to



AZ399400/c  
 LOCUS 1M0165C13F 21 bp DNA GSS 03-OCT-2000  
 DEFINITION clone UUGC1M0165C13 F, DNA sequence.  
 ACCESSION AZ399400  
 VERSION AZ399400.1 GI:10514474  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 21)  
 REFERENCE Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
 and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0165 row: C column: 13  
 Seq primer: CGTGTAAACGACGGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 21.  
 Location/Qualifiers  
 1..21  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0165C13"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptored DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi14732114|gbAF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptored mouse DNA was annealed to  
 adaptored vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 5 a 11 c 0 g 5 t  
 ORIGIN

Query Match 57.0%; Score 11.4; DB 243; Length 21;  
 Best Local Similarity 92.3%; Pred. No. 1.6e+05;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gttaggatggtg 14  
 |||||  
 Db 14 GTTAGGATGATG 2

RESULT 14  
 AZ786429  
 LOCUS 2M0031L23R 21 bp DNA GSS 16-FEB-2001  
 DEFINITION clone UUGC2M0031L23 R, DNA sequence.  
 ACCESSION AZ786429  
 VERSION AZ786429.1 GI:12924180  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 21)  
 REFERENCE Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
 and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0031 row: L column: 23  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 21.  
 Location/Qualifiers  
 1..21  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0031L23"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptored DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi14732114|gbAF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptored mouse DNA was annealed to  
 adaptored vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 2 a 1 c 13 g 5 t  
 ORIGIN

Query Match 57.0%; Score 11.4; DB 249; Length 21;  
 Best Local Similarity 92.3%; Pred. No. 1.6e+05;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gttaggatggtg 14  
 |||||  
 Db 8 GTTGGGATGGTG 20

RESULT 15

AZ822907 38 bp DNA GSS 20-FEB-2001  
LOCUS 2M0096M03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC2M0096M03 R, DNA sequence.

ACCESSION AZ822907

VERSION AZ822907.1 GI:12992815

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
AUTHORS 1 (bases 1 to 38)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.  
and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL plasmid inserts

COMMENT Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert length: 10000 Std Error: 0.00

Plate: 0096 row: M column: 03

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 38.

FEATURES Location/Qualifiers

source

1..38  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0096M03"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (g114732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 10 a 8 c 11 g 9 t  
ORIGIN

Query Match 57.0%; Score 11.4; DB 250; Length 38;  
Best Local Similarity 92.3%; Pred. No. 1.6e+05;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggttagggatggt 13

|||||

Db 20 GGTGCGGGATGCT 32







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:03:50 ; Search time 417.38 Seconds  
(without alignments)  
9.071 Million cell updates/sec

```

Title: US-09-757-100B-14
Perfect score: 20
Sequence: 1 qgttaqgqatggtgccatca 20

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 460742

```
Minimum DB seq length: 0
Maximum DB seq length: 50
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ins/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ins/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ins/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ins/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ins/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ins/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	20	100.0	20	3	US-09-377-310-14		Sequence 14, Appl
2	15	75.0	15	3	US-09-377-310-34		Sequence 34, Appl
3	14.2	71.0	34	4	US-09-004-731-58		Sequence 58, Appl
4	14.2	71.0	34	1	US-08-749-699-58		Sequence 58, Appl
C 5	13.8	69.0	24	4	US-08-460-344-18		Sequence 18, Appl
C 6	13.8	69.0	24	1	US-08-133-598A-18		Sequence 18, Appl
C 7	13.8	69.0	24	1	US-08-886-999-18		Sequence 18, Appl
C 8	13.8	69.0	24	5	PCI-US93-05065-18		Sequence 18, Appl
9	13.8	69.0	30	1	US-08-219-012-36		Sequence 36, Appl
10	13.8	69.0	30	4	US-08-687-421-224		Sequence 224, App
C 11	13.4	67.0	18	3	US-09-156-807-11		Sequence 11, Appl
C 12	13.4	67.0	18	3	US-09-156-807-12		Sequence 12, Appl
C 13	12.6	63.0	37	3	US-08-961-083-339		Sequence 339, App
C 14	12.6	63.0	50	6	5240845-45		Patent No. 5240845
15	12.2	61.0	20	3	US-09-418-641-75		Sequence 75, Appl
16	12	60.0	26	1	US-08-429-121B-87		Sequence 87, Appl
17	12	60.0	26	2	US-09-003-067-87		Sequence 87, Appl
C 18	11.8	59.0	26	1	US-08-429-121B-5		Sequence 5, Appl
C 19	11.8	59.0	26	1	US-08-429-121B-27		Sequence 27, Appl
C 20	11.8	59.0	26	1	US-08-429-121B-55		Sequence 55, Appl
C 21	11.8	59.0	26	1	US-08-429-121B-65		Sequence 65, Appl
C 22	11.8	59.0	26	1	US-08-429-121B-67		Sequence 67, Appl
C 23	11.8	59.0	26	1	US-08-429-121B-81		Sequence 81, Appl
C 24	11.8	59.0	26	1	US-08-429-121B-83		Sequence 83, Appl
C 25	11.8	59.0	26	1	US-08-429-121B-85		Sequence 85, Appl
C 26	11.8	59.0	26	2	US-09-003-067-5		Sequence 5, Appl
C 27	11.8	59.0	26	2	US-09-003-067-27		Sequence 27, Appl

## ALIGNMENTS

```

RESULT      1
US-09-377-310-14
; Sequence 14, Application US/09377310B
; Patent No. 6133031
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Gaarde, William A.
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase
; TITLE OF INVENTION: Expression
; FILE REFERENCE: ISPH-0389
; CURRENT APPLICATION NUMBER: US/09/377,310B
; CURRENT FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; -09-377-310-14
; OTHER INFORMATION: antisense sequence
US-09-377-310-14

```

```

Query Match      100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

r      1  ggtagggatggtgcgcgtca 20
      |||||
      1  ggtagggatggtgcgcgtca 20

```

```

RESULT      2
US-09-377-310-34
; Sequence 34, Application US/09377310B
; Patent No. 6133031
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Gaarde, William A.
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase
; TITLE OF INVENTION: Expression
; FILE REFERENCE: ISPH-0389
; CURRENT APPLICATION NUMBER: US/09/377,310B
; CURRENT FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Artificial Sequence

```

; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-34

Query Match 75.0%; Score 15; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ttagggatggtgcgcg 17  
Db 1 ttagggatggtgcgcg 15

## RESULT 3

US-09-004-731-58  
; Sequence 58, Application US/09004731  
; Patent No. 6177258  
; GENERAL INFORMATION:  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Stiegler, Gary  
; APPLICANT: Gaines, Patrick J.  
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES AND USES THEREOF  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/004,731  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA: US/08/749,699  
; APPLICATION NUMBER: US/08/749,699  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2618-25-C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 34 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (primer)  
US-09-004-731-58

Query Match 71.0%; Score 14.2; DB 4; Length 34;  
Best Local Similarity 84.2%; Pred. No. 71;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttaggatggtgcgcgta 20  
Db 13 GTTAGCGATGTCCTCCGTC 31

## RESULT 4

US-08-749-699-58  
; Sequence 58, Application US/08749699

; Patent No. 6210920  
; GENERAL INFORMATION:  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Stiegler, Gary  
; APPLICANT: Gaines, Patrick J.  
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES AND USES THEREOF  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/749,699  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2618-25-C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 34 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (primer)  
US-08-749-699-58

Query Match 71.0%; Score 14.2; DB 4; Length 34;  
Best Local Similarity 84.2%; Pred. No. 71;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttaggatggtgcgcgta 20  
Db 13 GTTAGCGATGTCCTCCGTC 31

## RESULT 5

US-08-460-344-18/c  
; Sequence 18, Application US/08460344  
; Patent No. 5654418  
; GENERAL INFORMATION:  
; APPLICANT: SHEINNESS, Diana K.  
; APPLICANT: ADAMS, Trevor H.  
; APPLICANT: STAMM, Michael R.  
; APPLICANT: CANGELOSI, Gerard A.  
; APPLICANT: BRITSCHGL, Theresa B.  
; APPLICANT: DIX, Connie K.  
; TITLE OF INVENTION: METHODS AND DIAGNOSTIC KITS USEFUL FOR  
; TITLE OF INVENTION: DETECTING MICROORGANISMS ASSOCIATED WITH VAGINAL  
; TITLE OF INVENTION: INFECTIONS  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: Steuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/460,344  
;; FILING DATE: 02-JUN-1995  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/133,598  
;; FILING DATE: 08-OCT-1993  
;; APPLICATION NUMBER: US 07/896,094  
;; FILING DATE: 29-MAY-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/600,334  
;; FILING DATE: 19-OCT-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Garrett-Wackowski, Eugenia  
;; REGISTRATION NUMBER: 37,330  
;; REFERENCE/DOCKET NUMBER: 11652-73-2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 543-9600  
;; TELEFAX: (415) 543-5043  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 24 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 1..24  
;; OTHER INFORMATION: /standard\_name= "GV003"  
US-08-460-344-18

Query Match 69.0%; Score 13.8; DB 1; Length 24;  
Best Local Similarity 88.2%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 ttagggtggtgcgcgc 19  
|| ||||| |||||  
Db 18 TTTGGGATGGAGCCGTC 2

RESULT 6  
US-08-133-598A-18/c  
; Sequence 18, Application US/08133598A  
; Patent No. 5700636  
; GENERAL INFORMATION:  
; APPLICANT: SHEINESS, Diana K.  
; APPLICANT: ADAMS, Trevor H.  
; APPLICANT: STAMM, Michael R.  
; APPLICANT: CANGELOSI, Gerard A.  
; APPLICANT: BRITSCHGI, Theresa B.  
; APPLICANT: DIX, Connie K.  
; TITLE OF INVENTION: METHODS AND DIAGNOSTIC KITS USEFUL FOR  
; TITLE OF INVENTION: DETECTING MICROORGANISMS ASSOCIATED WITH VAGINAL  
; TITLE OF INVENTION: INFECTIONS  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: Steuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/133,598A  
;; FILING DATE: 08-OCT-1993  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/896,094  
;; FILING DATE: 29-MAY-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/600,334  
;; FILING DATE: 19-OCT-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Garrett-Wackowski, Eugenia  
;; REGISTRATION NUMBER: 37,330  
;; REFERENCE/DOCKET NUMBER: 11652-73-2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 543-9600  
;; TELEFAX: (415) 543-5043  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 24 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 1..24  
;; OTHER INFORMATION: /standard\_name= "GV003"  
US-08-133-598A-18

Query Match 69.0%; Score 13.8; DB 1; Length 24;  
Best Local Similarity 88.2%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 ttagggtggtgcgcgc 19  
|| ||||| |||||  
Db 18 TTTGGGATGGAGCCGTC 2

RESULT 7  
US-08-886-999-18/c  
; Sequence 18, Application US/0886999  
; Patent No. 5776694  
; GENERAL INFORMATION:  
; APPLICANT: SHEINESS, Diana K.  
; APPLICANT: ADAMS, Trevor H.  
; APPLICANT: STAMM, Michael R.  
; APPLICANT: CANGELOSI, Gerard A.  
; APPLICANT: BRITSCHGI, Theresa B.  
; APPLICANT: DIX, Connie K.  
; TITLE OF INVENTION: METHODS AND DIAGNOSTIC KITS USEFUL FOR  
; TITLE OF INVENTION: DETECTING MICROORGANISMS ASSOCIATED WITH VAGINAL  
; TITLE OF INVENTION: INFECTIONS  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: Steuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/886,999  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/458,319

;; FILING DATE: 02-JUN-1995.  
;; APPLICATION NUMBER: US 08/133,598  
;; FILING DATE: 08-OCT-1993  
;; APPLICATION NUMBER: US 07/896,094  
;; FILING DATE: 29-MAY-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/600,334  
;; FILING DATE: 19-OCT-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Garrett-Wackowski, Eugenia  
;; REGISTRATION NUMBER: 37,330  
;; REFERENCE/DOCKET NUMBER: 11652-73-2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 543-9600  
;; TELEFAX: (415) 543-5043  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 24 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 1..24  
;; OTHER INFORMATION: /standard\_name= "GV003"  
US-08-886-999-18

Query Match 69.0%; Score 13.8; DB 1; Length 24;  
Best Local Similarity 88.2%; Pred. NO. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ttaggatggtgcgc 19  
|| ||||| |||||  
Db 18 TTTGGGATGGAGCCGTC 2

RESULT 8  
PCT-US93-05085-18/c  
;; GENERAL INFORMATION:  
;; APPLICANT: MICROPROBE CORPORATION  
;; TITLE OF INVENTION: METHODS AND PHARMACEUTICAL KITS USEFUL  
;; TITLE OF INVENTION: FOR DETECTING MICROORGANISMS ASSOCIATED WITH VAGINAL  
;; NUMBER OF SEQUENCES: 57  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend Khourie and Crew  
;; STREET: Steuart Street Tower, One Market Plaza  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: US  
;; ZIP: 94105-1493  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/05085  
;; FILING DATE: 28-MAY-1993  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Weber, Ellen L.  
;; REGISTRATION NUMBER: 32,762  
;; REFERENCE/DOCKET NUMBER: 11652-73-1PC  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 543-9600  
;; TELEFAX: (415) 543-5043  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 24 base pairs

;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
PCT-US93-05085-18

Query Match 69.0%; Score 13.8; DB 5; Length 24;  
Best Local Similarity 88.2%; Pred. NO. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ttaggatggtgcgc 19  
|| ||||| |||||  
Db 18 TTTGGGATGGAGCCGTC 2

RESULT 9  
US-08-219-012-36  
;; Sequence 36, Application US/08219012  
;; Patent No. 5543293  
;; GENERAL INFORMATION:  
;; APPLICANT: Larry Gold  
;; APPLICANT: Diane Tasset  
;; TITLE OF INVENTION: Ligands of Thrombin  
;; NUMBER OF SEQUENCES: 92  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Beaton & Swanson, P.C.  
;; STREET: 4582 South Ulster Street Parkway, Suite #  
;; CITY: Denver  
;; STATE: Colorado  
;; COUNTRY: USA  
;; ZIP: 80237  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
;; COMPUTER: IBM compatible  
;; OPERATING SYSTEM: MS-DOS  
;; SOFTWARE: WordPerfect 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/219,012  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA: none  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Barry J. Swanson  
;; REGISTRATION NUMBER: 33,215  
;; REFERENCE/DOCKET NUMBER:  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (303) 850-9900  
;; TELEFAX: (303) 850-9401  
;; INFORMATION FOR SEQ ID NO: 36:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 30 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
US-08-219-012-36

Query Match 69.0%; Score 13.8; DB 1; Length 30;  
Best Local Similarity 88.2%; Pred. NO. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggttagggatggtgcgc 17  
|| ||||| |||||  
Db 11 GGTtagggatggtgcgc 27

RESULT 10  
US-08-687-421-224  
;; Sequence 224, Application US/08687421  
;; Patent No. 6177557

```

; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Janjic, Nebojsa
; APPLICANT: Tasset, Diane
; TITLE OF INVENTION: HIGH-AFFINITY LIGANDS OF BASIC
; TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR AND
; TITLE OF INVENTION: THROMBIN
; NUMBER OF SEQUENCES: 445
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,421
; FILING DATE: 08-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/195,005
; FILING DATE: 10-FEBRUARY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE: 22-APRIL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/219,012
; FILING DATE: 28-MARCH-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,333
; FILING DATE: 11-NOVEMBER-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: 10-JUNE-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: 11-JUNE-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX07/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3433
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 224:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-687-421-224

Query Match 69.08; Score 13.8; DB 4; Len
Best Local Similarity 88.24; Pred. No. 1.1e+02;
Matches '15; Conservative 0; Mismatches 2; I

QY 1 ggttagggatgggtgccg 17
Db 11 GGTTAGGGATGGGTAGCG 27

RESULT 11
US-09-156-807-11/c
; Sequence 11, Application US/09156807
; Patent No. 6030786
; GENERAL INFORMATION:
; APPLICANT: Cowsett, Lex M.

```

```

RESULT 15
US-09-418-641-75
; Sequence 75, Application US/09418641A
; Patent No. 6124133
; GENERAL INFORMATION:
; APPLICANT: Jennifer K. Taylor
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF FRA-1 EXPRESSION
; FILE REFERENCE: RTS-0105
; CURRENT APPLICATION NUMBER: US/09/418,641A
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 75
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
;

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:48 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-15  
Perfect score: 20  
Sequence: 1 tgttggtttccaatcgacc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_ba3:\*
- 4: gb\_in1:\*
- 5: gb\_in2:\*
- 6: gb\_in3:\*
- 7: gb\_om:\*
- 8: gb\_ov:\*
- 9: gb\_pat1:\*
- 10: gb\_pat2:\*
- 11: gb\_ph:\*
- 12: gb\_pl1:\*
- 13: gb\_pl2:\*
- 14: gb\_pl3:\*
- 15: gb\_pl4:\*
- 16: em\_ba1:\*
- 17: em\_ba2:\*
- 18: em\_fun:\*
- 19: em\_htgo\_hum:\*
- 20: em\_htgo\_inv:\*
- 21: em\_htgo\_rod:\*
- 22: em\_htg\_hum1:\*
- 23: em\_htg\_hum2:\*
- 24: em\_htg\_hum3:\*
- 25: em\_htg\_hum4:\*
- 26: em\_htg\_hum5:\*
- 27: em\_htg\_hum6:\*
- 28: em\_htg\_hum7:\*
- 29: em\_htg\_hum8:\*
- 30: em\_htg\_inv1:\*
- 31: em\_htg\_inv2:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_rod:\*
- 34: em\_hum1:\*
- 35: em\_hum2:\*
- 36: em\_hum3:\*
- 37: em\_hum4:\*
- 38: em\_hum5:\*
- 39: em\_hum6:\*
- 40: em\_hum7:\*
- 41: em\_in:\*
- 42: em\_om:\*
- 43: em\_or:\*

- 44: em\_ov:\*
- 45: em\_pat:\*
- 46: em\_ph:\*
- 47: em\_pl:\*
- 48: em\_ro:\*
- 49: em\_sts:\*
- 50: em\_sy:\*
- 51: em\_un:\*
- 52: em\_vi:\*
- 53: gb\_sts1:\*
- 54: gb\_sts2:\*
- 55: gb\_sts3:\*
- 56: gb\_sy:\*
- 57: gb\_un:\*
- 58: gb\_vil:\*
- 59: gb\_vil2:\*
- 60: gb\_htg1:\*
- 61: gb\_htg2:\*
- 62: gb\_htg3:\*
- 63: gb\_htg4:\*
- 64: gb\_htg5:\*
- 65: gb\_htg6:\*
- 66: gb\_htg7:\*
- 67: gb\_htg8:\*
- 68: gb\_htg9:\*
- 69: gb\_htg10:\*
- 70: gb\_htg11:\*
- 71: gb\_htg12:\*
- 72: gb\_htg13:\*
- 73: gb\_htg14:\*
- 74: gb\_htg15:\*
- 75: gb\_htg16:\*
- 76: gb\_htg17:\*
- 77: gb\_htg18:\*
- 78: gb\_htg19:\*
- 79: gb\_htg20:\*
- 80: gb\_htg21:\*
- 81: gb\_htg22:\*
- 82: gb\_htg23:\*
- 83: gb\_htg24:\*
- 84: gb\_htg25:\*
- 85: gb\_pr1:\*
- 86: gb\_pr2:\*
- 87: gb\_pr3:\*
- 88: gb\_pr4:\*
- 89: gb\_pr5:\*
- 90: gb\_pr6:\*
- 91: gb\_pr7:\*
- 92: gb\_pr8:\*
- 93: gb\_pr9:\*
- 94: gb\_ro1:\*
- 95: gb\_ro2:\*
- 96: gb\_in4:\*
- 97: gb\_pr10:\*
- 98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
C 1	13.6	68.0	36	10	I09246	I09246 Sequence 12
C 2	13.6	68.0	37	10	I09244	I09244 Sequence 10
C 3	13.6	68.0	38	10	I09247	I09247 Sequence 13
C 4	13.4	67.0	30	9	AR109705	AR109705 Sequence
C 5	13.2	66.0	26	9	A56687	A56687 Sequence 6
C 6	13	65.0	19	10	E05267	E05267 Primer for
C 7	12.4	62.0	21	9	AX023416	AX023416 Sequence
C 8	12.4	62.0	24	9	AX052544	AX052544 Sequence

```

c 9      12      60.0      26 10 E32827 Novel compo
c 10     12      60.0      30 10 I69336 Sequence 8
c 11     12      60.0      42 9  AX019232 Sequence 6
c 12     12      60.0      48 10 I08124 Sequence 6
c 13     11.8     59.0      24 9  A43528 Sequence 8
c 14     11.8     59.0      24 9  AR052465 Sequence
c 15     11.8     59.0      35 9  AR011952 Sequence
c 16     11.8     59.0      35 9  AR093800 Sequence
c 17     11.6     58.0      27 9  AX023890 Sequence
c 18     11.6     58.0      27 95 S59402 TCR beta -T
c 19     11.6     58.0      30 9  A21431
c 20     11.6     58.0      30 9  A43129
c 21     11.6     58.0      34 9  AR091842 Sequence
c 22     11.6     58.0      41 9  AR109138 Sequence
c 23     11.6     58.0      50 10 E25881 Blastocyst
c 24     11.4     57.0      18 9  A26514 PT-26I muta
c 25     11.4     57.0      21 10 AX097321 Sequence
c 26     11.4     57.0      28 9  AR090859 Sequence
c 27     11.4     57.0      29 10 AX099592 Sequence
c 28     11.2     56.0      21 10 AX082611 Sequence
c 29     11.2     56.0      21 10 AX082613 Sequence
c 30     11.2     56.0      21 10 AX082616 Sequence
c 31     11.2     56.0      21 10 AX082618 Sequence
c 32     11.2     56.0      25 9  AR094119 Sequence
c 33     11.2     56.0      26 9  AR038156 Sequence
c 34     11.2     56.0      28 9  A14501 Group 8 Hyb
c 35     11.2     56.0      28 9  A14502 Group 8 bis
c 36     11.2     56.0      28 9  A14503 Group 8 ter
c 37     11.2     56.0      28 9  A14504 Group 8 qua
c 38     11.2     56.0      28 9  A14559 Hybridizati
c 39     11.2     56.0      30 9  A21427
c 40     11.2     56.0      32 10 AX085274
c 41     11.2     56.0      34 9  AR012356 Sequence
c 42     11.2     56.0      34 9  AX006548 Sequence
c 43     11.2     56.0      34 9  AX006652 Sequence
c 44     11.2     56.0      34 9  AX030256 Sequence
c 45     11.2     56.0      34 9  AX049934 Sequence

```

## ALIGNMENTS

```

RESULT 1
I09246/c 109246 Sequence 12 from Patent WO 8901940. PAT 02-DEC-1994
LOCUS I09246 36 bp
DEFINITION I09246
ACCESSION I09246
VERSION I09246.1 GI:588030
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 36)
AUTHORS Fisher, R.A., Gilbert, W., Sato, V.L., Flavell, R.A., Maraganore, J.M.
TITLE DNA SEQUENCES, RECOMBINANT DNA MOLECULES AND PROCESSES FOR
PRODUCING SOLUBLE T4 PROTEINS
JOURNAL Patent: WO 8901940-A 12 09-MAR-1989;
FEATURES
Location/Qualifiers
source 1..36
BASE COUNT 9 a 9 c 10 g 8 t
ORIGIN

Query Match 68.0%; Score 13.6; DB 10; Length 36;
Best Local Similarity 80.0%; Pred. No. 8.1e+03;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tgttggtttcccaatcgacc 20
||||| ||||| |||||
Db 28 TGTGGATTCCAGCAGGACC 9

RESULT 4
I09247/c 109247 Sequence 13 from Patent WO 8901940. PAT 02-DEC-1994
LOCUS I09247 38 bp
DEFINITION I09247
ACCESSION I09247
VERSION I09247.1 GI:588031
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 38)
AUTHORS Fisher, R.A., Gilbert, W., Sato, V.L., Flavell, R.A., Maraganore, J.M.
TITLE DNA SEQUENCES, RECOMBINANT DNA MOLECULES AND PROCESSES FOR
PRODUCING SOLUBLE T4 PROTEINS
JOURNAL Patent: WO 8901940-A 13 09-MAR-1989;
FEATURES
Location/Qualifiers
source 1..38
BASE COUNT 8 a 9 c 12 g 9 t
ORIGIN

Query Match 68.0%; Score 13.6; DB 10; Length 38;
Best Local Similarity 80.0%; Pred. No. 8.1e+03;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tgttggtttcccaatcgacc 20
||||| ||||| |||||
Db 15 TGTGGATTCCAGCAGGACC 34

RESULT 4
I09247/c 109247 Sequence 129 from patent US 6114139. PAT 14-FEB-2001
LOCUS I09247 30 bp
DEFINITION I09247
ACCESSION I09247
VERSION I09247.1 GI:12825981
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 30)
AUTHORS Fisher, R.A., Gilbert, W., Sato, V.L., Flavell, R.A., Maraganore, J.M.
TITLE DNA SEQUENCES, RECOMBINANT DNA MOLECULES AND PROCESSES FOR
PRODUCING SOLUBLE T4 PROTEINS
JOURNAL Patent: WO 8901940-A 13 09-MAR-1989;
FEATURES
Location/Qualifiers
source 1..30
BASE COUNT 8 a 9 c 12 g 9 t
ORIGIN

Query Match 68.0%; Score 13.6; DB 10; Length 38;
Best Local Similarity 80.0%; Pred. No. 8.1e+03;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tgttggtttcccaatcgacc 20
||||| ||||| |||||
Db 15 TGTGGATTCCAGCAGGACC 34

```

ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Hinuma,S., Hosoya,M., Fujii,R., Ohtaki,T., Fukusumi,S. and Ohgi,K.  
TITLE G-protein coupled receptor protein and a DNA encoding the receptor  
JOURNAL Patent: US 6114139-A 129 05-SEP-2000;  
FEATURES  
Location/Qualifiers  
1..30  
/organism="unknown"  
BASE COUNT 6 a 9 c 3 g 12 t  
ORIGIN  
|||||  
14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Query Match 67.0%; Score 13.4; DB 9; Length 30;  
Best Local Similarity 93.3%; Pred. No. 1.1e+04;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 tgggtgttccaatc 15  
Db 10 TGTGTTTCCAAC 24  
RESULT 5  
A56687/c A56687 26 bp DNA PAT 03-MAR-1998  
DEFINITION Sequence 6 from Patent EP0733708.  
ACCESSION A56687  
VERSION A56687.1 GI:3712729  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS Daban,M., Espuna,E., Medrano,A., Querol and Enrique.  
TITLE Transferrin-binding protein 1 (Tbpl) gene of Actinobacillus  
pleuropneumoniae, its use in vaccines for pleuropneumonia and as  
diagnostic reagents  
JOURNAL Patent: EP 0733708-A 6 25-SEP-1996;  
HIPRA S A LAB (ES)  
COMMENT Other publication ES 2092960 961201  
Other publication CA 2172443 960925.  
FEATURES  
source  
1..26  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 9 a 6 c 9 g 2 t  
ORIGIN  
Query Match 66.0%; Score 13.2; DB 9; Length 26;  
Best Local Similarity 83.3%; Pred. No. 1.4e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 2 gttggttccaatcgag 19  
Db 23 GTTGTTCACCTCTAC 6  
RESULT 6  
E05267/c E05267 19 bp DNA PAT 29-SEP-1997  
LOCUS Primer for amplifying Human papillomavirus Sapporo gene.  
DEFINITION E05267  
ACCESSION E05267  
VERSION E05267.1 GI:2173457  
KEYWORDS JP 1993192200-A/37.  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 19)  
AUTHORS Okazawa,K., Shimada,M., Katou,I., Fukushima,M. and Fujinaga,K.  
TITLE DETECTION OF HUMAN PAPILLOMA VIRUS  
JOURNAL Patent: JP 1993192200-A 37 03-AUG-1993;  
TAKARA SHUZO CO LTD

COMMENT OS Artificial gene  
OC Artificial sequence; Genes.  
PN JP 1993192200-A/37  
PD 03-AUG-1993  
PF 19-AUG-1991 JP 1991230839  
PR 20-AUG-1990 JP 90P 217067  
PI OKAZAWA KAZUHIDE, SHIMADA MASAMITSU, KATOU IKUNOSHIN, PI  
FUKUSHIMA MICHIO,  
FUJINAGA KEI  
PC C1201/70,C1201/68;  
CC strandedness: Single;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
FEATURES  
Location/Qualifiers  
1..19  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 7 a 4 c 5 g 3 t  
ORIGIN  
Query Match 65.0%; Score 13; DB 10; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.8e+04;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 tgttggtttccaa 13  
Db 14 TGTGTTTCCA 2  
RESULT 7  
AX023416 21 bp DNA PAT 15-SEP-2000  
LOCUS Sequence 31 from Patent WO0014217.  
DEFINITION AX023416  
ACCESSION AX023416  
VERSION AX023416.1 GI:10183816  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Lipford,G.B., Heeg,K. and Wagner,H.  
TITLE G-motif oligonucleotides and uses thereof  
JOURNAL Patent: WO 0014217-A 31 16-MAR-2000;  
LIPFORD GRAYSON B (DE) ; HEEG KLAUS (DE) ; WAGNER HERMANN (DE) ;  
CPG IMMUNOPHARMACEUTICALS GMBH (DE)  
FEATURES  
Location/Qualifiers  
1..21  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="synthetic, no natural origin"  
BASE COUNT 2 a 5 c 3 g 11 t  
ORIGIN  
Query Match 62.0%; Score 12.4; DB 9; Length 21;  
Best Local Similarity 92.9%; Pred. No. 4.1e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 tgttggtttccaa 14  
Db 8 TGTGTTTCCAT 21  
RESULT 8  
AX052544 24 bp DNA PAT 12-JAN-2001  
LOCUS Sequence 6 from Patent WO0071668.  
DEFINITION AX052544  
ACCESSION AX052544  
VERSION AX052544.1 GI:12226774  
KEYWORDS  
SOURCE synthetic construct.

RESULT	10
I69336/c	
LOCUS	I69336 30 bp DNA PAT 04-FEB-1998
DEFINITION	Sequence 8 from patent US 5677172.
ACCESSION	I69336
VERSION	I69336.1 GI:2831458
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 30)
TITLE	Makarow,M.
JOURNAL	Method for production of proteins in yeast
FEATURES	Patent: US 5677172-A 8 14-OCT-1997; Location/Qualifiers 1..30 /organism="unknown"
BASE COUNT	6 a 9 c 8 g 7 t
ORIGIN	
Query Match	60.0%; Score 12; DB 10; Length 30;
Best Local Similarity	75.0%; Pred. No. 6.9e+04;
Matches	15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy	1 ttgttggtttccaatcgacc 20 
Db	24 TGTGGTTCACAAGGTACC 5
RESULT	11
AX019232/c	
LOCUS	AX019232 42 bp DNA PAT 07-SEP-2000
DEFINITION	Sequence 12 from Patent WO9941390.
ACCESSION	AX019232
VERSION	AX019232.1 GI:10043261
KEYWORDS	
SOURCE	synthetic construct. synthetic construct artificial sequence.
REFERENCE	1 (bases 1 to 42)
AUTHORS	Huels,C., Muellner,S., Boehnisch,B. and Gallert,K.C.
TITLE	Expression vector for the production of dead proteins
JOURNAL	Patent: WO 9941390-A 12 19-AUG-1999; HUELS CHRISTOPH (DE); MUELLNER STEFAN (DE); AVENTIS RES & TECH GMBH & CO (DE); BOEHNISCH BRITTA (DE); GALLERT KARL CHRISTIAN (DE)
FEATURES	Location/Qualifiers 1..42 /organism="synthetic construct" /db_xref="taxon:32630" /note="pi35-Nr5C"
BASE COUNT	12 a 13 c 6 g 11 t
ORIGIN	
Query Match	60.0%; Score 12; DB 9; Length 42;
Best Local Similarity	75.0%; Pred. No. 6.8e+04;
Matches	15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy	1 ttgttggtttccaatcgacc 20 
Db	32 TGATGATGCCCATCGGATC 13
RESULT	12
I08124/c	
LOCUS	I08124 48 bp PAT 02-DEC-1994
DEFINITION	Sequence 6 from Patent EP 0334004.
ACCESSION	I08124
VERSION	I08124.1 GI:589165
KEYWORDS	
SOURCE	Unknown.

```
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 48)
AUTHORS Smith,J.A. and Lee,F-J.S.
TITLE Isolation, purification, characterization, cloning and sequencing
of N alpha-acetyltransferase
JOURNAL Patent: EP 0334004-A1 6 27-SEP-1989;
FEATURES Location/Qualifiers
source
1..48
/organism="unknown"
BASE COUNT 14 a 10 c 13 g 11 t
ORIGIN

Query Match 60.0%; Score 12; DB 10; Length 48;
Best Local Similarity 75.0%; Pred. No. 6.8e+04;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tgggtttccaatcgacc 20
Db 43 TGCTGCTTACCACATCTGACC 24

RESULT 13
LOCUS A43528 24 bp DNA PAT 06-MAR-1997
DEFINITION Sequence 8 from Patent WO9507981.
ACCESSION A43528
VERSION A43528.1 GI:2298716
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 24)
AUTHORS Schweighoffer,F. and Tocque,B.
TITLE GRB3-3 GENE, VARIANTS AND USES THEREOF
JOURNAL Patent: WO 9507981-A 8 23-MAR-1995;
COMMENT RHONE POULENC RORER SA (FR)
Other publication PL 313445 960708
Other publication NO 960965 960308
Other publication FI 961202 960314
Other publication CA 2169338 950323
Other publication ZA 9407059 950518
Other publication AU 6724794 950403
Other publication FR 2710074 950324.
FEATURES Location/Qualifiers
source
1..24
/organism="synthetic construct"
BASE COUNT 5 a 4 c 6 g 9 t
ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 24;
Best Local Similarity 86.7%; Pred. No. 9e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 tgggtttccaatcgga 18
Db 2 TCGTTTCCAAACGGA 16

RESULT 14
LOCUS AR052465 24 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 9 from patent US 5831048.
ACCESSION AR052465
VERSION AR052465.1 GI:5975829
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 24)
```

```
AUTHORS Schweighoffer,F. and Tocque ,B.
TITLE GRB3-3 CDNA and polypeptides
JOURNAL Patent: US 5831048-A 9 03-NOV-1998;
FEATURES Location/Qualifiers
source
1..24
/organism="unknown"
BASE COUNT 5 a 4 c 6 g 9 t
ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 24;
Best Local Similarity 86.7%; Pred. No. 9e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 tgggtttccaatcgga 18
Db 2 TCGTTTCCAAACGGA 16

RESULT 15
LOCUS AR011952 35 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 29 from patent US 5763177.
ACCESSION AR011952
VERSION AR011952.1 GI:3969942
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 35)
AUTHORS Gold,L., Willis,M., Koch,T., Ringquist,S., Jensen,K. and
Atkinson,B.
TITLE Systematic evolution of ligands by exponential enrichment:
photoselection of nucleic acid ligands and solution sele
JOURNAL Patent: US 5763177-A 29 09-JUN-1998;
FEATURES Location/Qualifiers
source
1..35
/organism="unknown"
BASE COUNT 8 a 8 c 6 g 13 t
ORIGIN

Query Match 59.0%; Score 11.8; DB 9; Length 35;
Best Local Similarity 86.7%; Pred. No. 9e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgggtttccaatc 15
Db 18 TCGTTTGTTCACATC 32

Search completed: October 2, 2001, 15:56:49
Job time: 14173 sec
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:02 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-15

Perfect score: 20

Sequence: 1 tttgtgttccaatcgacc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est25:\*
- 26: gb\_est26:\*
- 27: gb\_est27:\*
- 28: gb\_est28:\*
- 29: gb\_est29:\*
- 30: gb\_est30:\*
- 31: gb\_est31:\*
- 32: gb\_est32:\*
- 33: gb\_est33:\*
- 34: gb\_est34:\*
- 35: gb\_est35:\*
- 36: gb\_est36:\*
- 37: gb\_est37:\*
- 38: gb\_est38:\*
- 39: gb\_est39:\*
- 40: gb\_est40:\*
- 41: gb\_est41:\*
- 42: gb\_est42:\*
- 43: gb\_est43:\*
- 44: gb\_est44:\*
- 45: gb\_est45:\*
- 46: gb\_est46:\*
- 47: gb\_est47:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est33:\*
- 111: gb\_est34:\*
- 112: gb\_est35:\*
- 113: gb\_est36:\*
- 114: gb\_est37:\*
- 115: gb\_est38:\*
- 116: gb\_est39:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: em\_est100:\*  
170: em\_est101:\*  
171: em\_est102:\*  
172: em\_est103:\*  
173: em\_est104:\*  
174: em\_est105:\*  
175: em\_est106:\*  
176: em\_est107:\*  
177: em\_est108:\*  
178: em\_est109:\*  
179: em\_est110:\*  
180: em\_est111:\*  
181: em\_est112:\*  
182: em\_est113:\*  
183: em\_est114:\*  
184: em\_est115:\*  
185: em\_est116:\*  
186: em\_est117:\*  
187: em\_est118:\*  
188: em\_est119:\*  
189: em\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: em\_gss\_fun:\*  
194: em\_gss\_hum1:\*  
195: em\_gss\_hum2:\*  
196: em\_gss\_hum3:\*  
197: em\_gss\_hum4:\*  
198: em\_gss\_hum5:\*  
199: em\_gss\_hum6:\*  
200: em\_gss\_hum7:\*  
201: em\_gss\_hum8:\*  
202: em\_gss\_hum9:\*  
203: em\_gss\_inv1:\*  
204: em\_gss\_inv2:\*  
205: em\_gss\_inv3:\*  
206: em\_gss\_inv4:\*  
207: em\_gss\_inv5:\*  
208: em\_gss\_inv6:\*  
209: em\_gss\_inv7:\*  
210: em\_gss\_inv8:\*  
211: em\_gss\_inv9:\*  
212: em\_gss\_inv10:\*  
213: em\_gss\_inv11:\*  
214: em\_gss\_inv12:\*  
215: em\_gss\_inv13:\*  
216: em\_gss\_inv14:\*  
217: em\_gss\_inv15:\*  
218: em\_gss\_inv16:\*  
219: em\_gss\_inv17:\*  
220: em\_gss\_inv18:\*  
221: em\_gss\_inv19:\*  
222: em\_gss\_inv20:\*  
223: em\_gss\_inv21:\*  
224: em\_gss\_inv22:\*  
225: em\_gss\_inv23:\*  
226: em\_gss\_inv24:\*  
227: em\_gss\_inv25:\*  
228: em\_gss\_inv26:\*  
229: em\_gss\_inv27:\*  
230: em\_gss\_inv28:\*  
231: em\_gss\_inv29:\*  
232: em\_gss\_inv30:\*  
233: em\_gss\_inv31:\*  
234: em\_gss\_inv32:\*  
235: em\_gss\_inv33:\*  
236: em\_gss\_inv34:\*  
237: em\_gss\_inv35:\*  
238: em\_gss\_inv36:\*  
239: em\_gss\_inv37:\*  
240: em\_gss\_inv38:\*  
241: em\_gss\_inv39:\*  
242: em\_gss\_inv40:\*  
243: em\_gss\_inv41:\*  
244: em\_gss\_inv42:\*  
245: em\_gss\_inv43:\*  
246: em\_gss\_inv44:\*  
247: em\_gss\_inv45:\*  
248: em\_gss\_inv46:\*  
249: em\_gss\_inv47:\*  
250: em\_gss\_inv48:\*  
251: em\_gss\_inv49:\*  
252: em\_gss\_inv50:\*  
253: em\_gss\_inv51:\*  
254: em\_gss\_inv52:\*  
255: em\_gss\_inv53:\*  
256: em\_gss\_inv54:\*  
257: em\_gss\_inv55:\*  
258: em\_gss\_inv56:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	12.8	64.0	46	12	AA790759
C 2	12.2	61.0	29	250	AZ838877
C 3	12.2	61.0	30	249	AZ768472
C 4	12.0	60.0	28	242	AZ380625
C 5	12.0	60.0	37	258	TA379B06Q
C 6	12.0	60.0	40	245	AZ537263
C 7	12.0	60.0	43	13	AA895268
C 8	12.0	60.0	43	241	AZ319753
C 9	11.8	59.0	50	107	AU107355
C 10	11.6	58.0	19	241	AZ327390
C 11	11.4	57.0	46	23	AI663836
C 12	11.4	57.0	50	117	AW537238
C 13	11.2	56.0	40	10	AA652673
C 14	11.2	56.0	40	10	AA652673
C 15	11.2	56.0	43	250	AZ822833
C 16	11.1	55.0	31	1	AA043719
C 17	11.1	55.0	36	243	AZ427281
C 18	11.1	55.0	39	249	AZ762518
C 19	11.1	55.0	40	24	AI783759
C 20	11.1	55.0	43	189	T63717
C 21	11.1	55.0	50	24	AI750109
C 22	11.1	55.0	50	107	AU103004
C 23	11.1	55.0	50	107	AU105668
C 24	10.8	54.0	22	246	AZ623308
C 25	10.8	54.0	22	246	AZ623308
C 26	10.8	54.0	37	249	AZ803825
C 27	10.8	54.0	40	166	BE377955
C 28	10.8	54.0	44	244	AZ479665
C 29	10.8	54.0	45	249	AZ791420
C 30	10.8	54.0	46	1	AA037870
C 31	10.8	54.0	46	155	C02388
C 32	10.8	54.0	46	159	H98043
C 33	10.6	53.0	22	246	AZ601191
C 34	10.6	53.0	22	249	AZ787023
C 35	10.6	53.0	29	249	AZ779056
C 36	10.6	53.0	37	20	AI446151
C 37	10.6	53.0	42	249	AZ779815
C 38	10.6	53.0	45	244	AZ461904
C 39	10.6	53.0	45	250	AZ819815
C 40	10.6	53.0	47	258	TA220G01P
C 41	10.6	53.0	49	2	RA115149
C 42	10.6	53.0	50	107	AU108011
C 43	10.6	53.0	50	249	AZ794281
C 44	10.4	52.0	23	242	AZ350054
C 45	10.4	52.0	28	241	AZ324675

## ALIGNMENTS

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	12.8	64.0	46	12	AA790759
C 2	12.2	61.0	29	250	AZ838877
C 3	12.2	61.0	30	249	AZ768472
C 4	12.0	60.0	28	242	AZ380625
C 5	12.0	60.0	37	258	TA379B06Q
C 6	12.0	60.0	40	245	AZ537263
C 7	12.0	60.0	43	13	AA895268
C 8	12.0	60.0	43	241	AZ319753
C 9	11.8	59.0	50	107	AU107355
C 10	11.6	58.0	19	241	AZ327390
C 11	11.4	57.0	46	23	AI663836
C 12	11.4	57.0	50	117	AW537238
C 13	11.2	56.0	40	10	AA652673
C 14	11.2	56.0	40	10	AA652673
C 15	11.2	56.0	43	250	AZ822833
C 16	11.1	55.0	31	1	AA043719
C 17	11.1	55.0	36	243	AZ427281
C 18	11.1	55.0	39	249	AZ762518
C 19	11.1	55.0	40	24	AI783759
C 20	11.1	55.0	43	189	T63717
C 21	11.1	55.0	50	24	AI750109
C 22	11.1	55.0	50	107	AU103004
C 23	11.1	55.0	50	107	AU105668
C 24	10.8	54.0	22	246	AZ623308
C 25	10.8	54.0	22	246	AZ623308
C 26	10.8	54.0	37	249	AZ803825
C 27	10.8	54.0	40	166	BE377955
C 28	10.8	54.0	44	244	AZ479665
C 29	10.8	54.0	45	249	AZ791420
C 30	10.8	54.0	46	1	AA037870
C 31	10.8	54.0	46	155	C02388
C 32	10.8	54.0	46	159	H98043
C 33	10.6	53.0	22	246	AZ601191
C 34	10.6	53.0	22	249	AZ787023
C 35	10.6	53.0	29	249	AZ779056
C 36	10.6	53.0	37	20	AI446151
C 37	10.6	53.0	42	249	AZ779815
C 38	10.6	53.0	45	244	AZ461904
C 39	10.6	53.0	45	250	AZ819815
C 40	10.6	53.0	47	258	TA220G01P
C 41	10.6	53.0	49	2	RA115149
C 42	10.6	53.0	50	107	AU108011
C 43	10.6	53.0	50	249	AZ794281
C 44	10.4	52.0	23	242	AZ350054
C 45	10.4	52.0	28	241	AZ324675

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	12.8	64.0	46	12	AA790759
C 2	12.2	61.0	29	250	AZ838877
C 3	12.2	61.0	30	249	AZ768472
C 4	12.0	60.0	28	242	AZ380625
C 5	12.0	60.0	37	258	TA379B06Q
C 6	12.0	60.0	40	245	AZ537263
C 7	12.0	60.0	43	13	AA895268
C 8	12.0	60.0	43	241	AZ319753
C 9	11.8	59.0	50	107	AU107355
C 10	11.6	58.0	19	241	AZ327390
C 11	11.4	57.0	46	23	AI663836
C 12	11.4	57.0	50	117	AW537238
C 13	11.2	56.0	40	10	AA652673
C 14	11.2	56.0	40	10	AA652673
C 15	11.2	56.0	43	250	AZ822833
C 16	11.1	55.0	31	1	AA043719
C 17	11.1	55.0	36	243	AZ427281
C 18	11.1	55.0	39	249	AZ762518
C 19	11.1	55.0	40	24	AI783759
C 20	11.1	55.0	43	189	T63717
C 21	11.1	55.0	50	24	AI750109
C 22	11.1	55.0	50	107	AU103004
C 23	11.1	55.0	50	107	AU105668
C 24	10.8	54.0	22	246	AZ623308
C 25	10.8	54.0	22	246	AZ623308
C 26	10.8	54.0	37	249	AZ803825
C 27	10.8	54.0	40	166	BE377955
C 28	10.8	54.0	44	244	AZ479665
C 29	10.8	54.0	45	249	AZ791420
C 30	10.8	54.0	46	1	AA037870
C 31	10.8	54.0	46	155	C02388
C 32	10.8	54.0	46	159	H98043
C 33	10.6	53.0	22	246	AZ601191
C 34	10.6	53.0	22	249	AZ787023
C 35	10.6	53.0	29	249	AZ779056
C 36	10.6	53.0	37	20	AI446151
C 37	10.6	53.0	42	249	AZ779815
C 38	10.6	53.0	45	244	AZ461904
C 39	10.6	53.0	45	250	AZ819815
C 40	10.6	53.0	47	258	TA220G01P
C 41	10.6	53.0	49	2	RA115149
C 42	10.6	53.0	50	107	AU108011
C 43	10.6	53.0	50	249	AZ794281
C 44	10.4	52.0	23	242	AZ350054
C 45	10.4	52.0	28	241	AZ324675

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	12.8	64.0	46	12	AA790759
C 2	12.2	61.0	29	250	AZ838877
C 3	12.2	61.0	30	249	AZ768472
C 4	12.0	60.0	28	242	AZ380625
C 5	12.0	60.0	37	258	TA379B06Q
C 6	12.0	60.0	40	245	AZ537263
C 7	12.0	60.0	43	13	AA895268
C 8	12.0	60.0	43	241	AZ319753
C 9	11.8	59.0	50	107	AU107355
C 10	11.6	58.0	19	241	AZ327390
C 11	11.4	57.0	46	23	AI663836
C 12	11.4	57.0	50	117	AW537238
C 13	11.2	56.0	40	10	AA652673
C 14	11.2	56.0	40	10	AA652673
C 15	11.2	56.0	43	250	AZ822833
C 16	11.1	55.0	31	1	AA043719
C 17	11.1	55.0	36	243	AZ427281
C 18	11.1	55.0	39	249	AZ762518
C 19	11.1	55.0	40	24	AI783759
C 20	11.1	55.0	43	189	T63717
C 21	11.1	55.0	50	24	AI750109
C 22	11.1	55.0	50	107	AU103004
C 23	11.1	55.0	50	107	AU105668
C 24	10.8	54.0	22	246	AZ623308
C 25	10.8	54.0	22	246	AZ623308
C 26	10.8	54.0	37	249	AZ803825
C 27	10.8	54.0	40	166	BE377955
C 28	10.8	54.0	44	244	AZ479665
C 29	10.8	54.0	45	249	AZ791420
C 30	10.8	54.0	46	1	AA037870
C 31	10.8	54.0	46	155	C02388
C 32	10.8	54.0	46	159	H98043
C 33	10.6	53.0	22	246	AZ601191
C 34	10.6	53.0	22	249	AZ787023
C 35	10.6	53.0	29	249	AZ779056
C 36	10.6	53.0	37	20	AI446151
C 37	10.6	53.0	42	249	AZ779815
C 38	10.6	53.0	45	244	AZ461904
C 39	10.6	53.0	45	250	AZ819815
C 40	10.6	53.0	47	258	TA220G01P
C 41	10.6	53.0	49	2	RA115149
C 42	10.6	53.0	50	107	AU108011
C 43	10.6	53.0	50	249	AZ794281
C 44	10.4	52.0	23	242	AZ350054
C 45	10.4	52.0	28	241	AZ324675

RESULT	1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
--------	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

## COMMENT

Contact: Robert B. Weiss  
University of Utah  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0134 row: E column: 10  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 29.

## FEATURES

source

1..29  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="UUCG2M0134E10"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil147321141gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

6 a 7 c 5 g 11 t

Query Match 61.0%; Score 12.2; DB 250; Length 29;  
Best Local Similarity 82.4%; Pred. No. 2 4e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttgggtttccaatcgac 19  
||| || ||| ||| ||| |||  
Db 29 TTGATTACCACTCGGAC 13

## RESULT 3

AZ768472 30 bp DNA 16-FEB-2001  
LOCUS  
DEFINITION 1M0568C10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0568C10 R, DNA sequence.  
ACCESSION AZ768472  
VERSION  
KEYWORDS  
SOURCE GSS.  
house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 30)  
REFERENCE  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

## JOURNAL COMMENT

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0568 row: C column: 10  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 30.

## FEATURES

source

1..30  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="UUGC1M0568C10"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil147321141gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 7 a 3 c 8 g 12 t

## ORIGIN

Query Match 61.0%; Score 12.2; DB 249; Length 30;  
Best Local Similarity 82.4%; Pred. No. 2 4e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggtttccaatcgga 18  
||| ||| ||| ||| ||| |||  
Db 3 GTTGGTTTACATTCGA 19

## RESULT 4

AZ380625/c 28 bp DNA 02-OCT-2000  
LOCUS  
DEFINITION 1M0136G10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0136G10 R, DNA sequence.  
ACCESSION AZ380625  
VERSION  
KEYWORDS  
SOURCE GSS.  
house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 28)  
REFERENCE  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL  
COMMENT

plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0136 row: G column: 10  
Seq primer: CACACGAGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 28.

FEATURES  
source

1..28  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0136G10"  
/clone.lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gii4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

9 a 3 c 9 g 7 t  
Query Match 60.0%; Score 12; DB 242; Length 28;  
Best Local Similarity 75.0%; Pred. No. 3e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tggtagtttccaatcgacc 20  
||||| ||||| ||||| |||||

Db 24 TGTAGCTTCCCATCTGAAC 5

RESULT  
TA379B06Q/C

LOCUS TA379B06Q 37 bp DNA GSS 13-DEC-2000  
DEFINITION T. brucei sheared genomic DNA clone 379b06, reverse sequence,  
genomic survey sequence.  
ACCESSION AL497861  
VERSION AL497861.1 GI:11873583  
KEYWORDS GSS.

SOURCE  
ORGANISM

Trypanosoma brucei.  
Trypanosoma.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 37)  
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,  
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,  
Melville, S.E., Rajandream, M.A. and Barrell, B.G.  
Direct Submission

## TITLE

## JOURNAL

## COMMENT

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
nh@sanger.ac.uk  
Constructed at the Institute for Genomic Research (TIGR),  
Rockville, MD. Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
to give a tight size distribution (4 kb). The v + i method used for the library construction is  
described in detail in Smith, H. and Venter, J.C. (Making small  
insert libraries for whole genome shotgun sequencing projects. In  
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999).  
Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available  
at http://www.sanger.ac.uk/Projects/T\_brucei/.

FEATURES  
source

1..37  
/organism="Trypanosoma brucei"  
/strain="TREU927"  
/db\_xref="taxon:5691"  
/clone="379b06"

BASE COUNT 12 a 10 c 6 g 9 t  
ORIGIN

Query Match 60.0%; Score 12; DB 258; Length 37;  
Best Local Similarity 75.0%; Pred. No. 3.2e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tggtagtttccaatcgacc 20  
||||| ||||| ||||| |||||

Db 21 TGTGATTTCCAGAGAAC 2

RESULT  
AZ537263

LOCUS AZ537263 40 bp DNA GSS 06-NOV-2000  
DEFINITION ASP-2P03160 Genetrap PC-3 Human Prostatic Carcinoma Library Homo  
sapiens genomic 5', DNA sequence.  
ACCESSION AZ537263  
VERSION AZ537263.1 GI:11114226  
KEYWORDS GSS.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS

Henkel, G., Liyanage, M., Pratt, E., Huang, D., Riley, M., Bernardino, A.,  
Durick, K. and Pollok, B.

TITLE  
JOURNAL

Exon-trap tags from a PC-3 GenomesScreen(TM) Library  
Unpublished (2000)

## COMMENT

Contact: Greg Henkel  
Gene Expression  
Aurora Biosciences Corp.  
11010 Torreyana Road, San Diego, CA 92121, USA  
Tel: 8584048436  
Fax: 8584046719  
Email: henkelg@aurorabio.com

Pools of cells were isolated from a GenomesScreen(TM) library. The  
library of cells was generated by retroviral integration of a gene  
tagging element consisting of: 1) A promoterless beta-lactamase  
preceded by a splice acceptor as a reporter for gene expression;  
2) A promoter driving neomycin resistance followed by a splice  
donor to trap downstream exons. 3' RACE from neomycin gene was  
performed using total RNA from isolated pools. Output was shotgun  
bacteriophage lambda and used to transform DH5-alpha competent  
bacteria. 5' ends of reported sequences were immediately preceded  
by splice donor from the trapping construct.

FEATURES  
source

Location/Qualifiers  
1..40  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
 /clone\_lib="Genetrap PC-3 Human Prostatic Carcinoma  
 Library"  
 /tissue\_type="Adenocarcinoma"  
 /cell\_type="Epithelial"  
 /note="Organ: PC-3"  
 RNA from genetrap pools; shotgun clone in pAmp-1 and used  
 to transform DH5-alpha competent bacteria."

BASE COUNT 12 a 12 c 4 g 12 t  
 ORIGIN

Query Match 60.0%; Score 12; DB 245; Length 40;  
 Best Local Similarity 75.0%; Pred. No. 3.3e+04;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttttggtttccaatcgacc 20  
 ||||| ||||| | |||  
 Db 5 TGTGTGATTTCCACATGAAC 24

RESULT 7  
 AA895268/c  
 LOCUS AA895268 43 bp mRNA EST 06-APR-1998  
 DEFINITION vx49c02.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone  
 IMAGE:1278530 5' similar to TR:002754 O02754 C/EBP ALPHA PROTEIN.  
 ; mRNA sequence.

ACCESSION AA895268  
 VERSION AA895268.1 GI:3031687  
 KEYWORDS EST.  
 SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 43)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:670330

Trace considered overall poor quality  
 Possible reversed clone: Similarity on wrong strand  
 Seq primer: -28m3 revl Et from Amersham  
 High quality sequence stop: 1.

FEATURES  
 source

Location/Qualifiers  
 1..43  
 /organism="Mus musculus"  
 /strain="C57BL/6 x CBA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1278530"  
 /clone\_lib="Stratagene mouse lung 937302"  
 /sex="female"  
 /tissue\_type="lung"  
 /dev\_host="6-8 month old"  
 /lab\_host="SOL (kanamycin resistant)"  
 /note="Organ: lung; Vector: pBluescript SK-; Site:1; EcorI  
 ; Site:2; XhoI; Cloned unidirectionally. Primer: Oligo  
 dT. 6-8 month old female lung and 1.5 year old male lung  
 were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP  
 XR Vector; -5' adaptor sequence: 5' GAATTCGGCAG 3' -3'

adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3'  
 BASE COUNT 12 a 12 c 14 g 5 t  
 ORIGIN

Query Match 60.0%; Score 12; DB 13; Length 43;  
 Best Local Similarity 75.0%; Pred. No. 3.3e+04;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttttggtttccaatcgacc 20  
 ||||| ||||| | |||  
 Db 32 TGTTCGGTTCCACGATACC 13

RESULT 8  
 A2319753  
 LOCUS A2319753 43 bp DNA GSS 29-SEP-2000  
 DEFINITION IM0039802R Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0039802 R, DNA sequence.

ACCESSION A2319753  
 VERSION A2319753.1 GI:10370842  
 KEYWORDS GSS  
 SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 43)

REFERENCE 1  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0039 row: B column: 02

Seq primer: CACACAGGAAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 43.

FEATURES  
 source

Location/Qualifiers  
 1..43  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0039802"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 [gi14732114|gb|AF129072.1], a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into



ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS 1 (bases 1 to 27)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL  
COMMENT Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0095 row: M column: 09  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 27.

FEATURES  
Location/Qualifiers  
1..27  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUCG2M0095M09"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g114732114 [gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 7 a 3 c 8 g 9 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 250; Length 27;  
Best Local Similarity 77.8%; Pred. No. 5e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ttgtggtttccatcgga 18  
||| ||||| |||||  
Db 1 TGGTGATTTCCTATGGGA 18

RESULT 12  
A1663836/c  
LOCUS A1663836 46 bp mRNA EST 10-MAY-1999  
DEFINITION uJ06e03.xl Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:1891132 3' similar to SW:ALBU\_MERUN O35090 SERUM ALBUMIN PRECURSOR. ; mRNA sequence.

ACCESSION A1663836  
VERSION A1663836.1 GI:4767419  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS 1 (bases 1 to 46)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL  
COMMENT Unpublished (1999)  
Other\_ESTs: uJ06e03.yl  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:975456

FEATURES  
Location/Qualifiers  
1..46  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1891132"  
/clone\_lib="Sugano mouse liver mlia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: liver; Vector: pME18S-FL3; Site:1: DraIII (CACGTGTC); Site:2: DraIII (CACCAGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGGCGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACGTGTC, 3' site CACCAGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

BASE COUNT 13 a 12 c 11 g 10 t  
ORIGIN

Query Match 57.0%; Score 11.4; DB 23; Length 46;  
Best Local Similarity 92.3%; Pred. No. 7.1e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 ttggtttccatc 15  
||| ||||| |||||  
Db 23 TTTGTTTCCATC 11

RESULT 13  
AW537238/c  
LOCUS AW537238 50 bp mRNA EST 31-AUG-2000  
DEFINITION G0114A01-3 Mouse E7.5 Embryonic Portion cDNA Library Mus musculus cDNA clone G0114A01 3', mRNA sequence.

ACCESSION AW537238  
VERSION AW537238.1 GI:7179655  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.



REFERENCE 1 (bases 1 to 50)  
 AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H., III, Becker,K.G. and Ko,M.S.H.  
 TITLE Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
 MEDLINE 20381348  
 COMMENT Contact: George J. Kargul  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdaelg@un.gcr.nia.nih.gov  
 Plate: G0114 row: A column: 01  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 50  
 POLYA-Yes.

FEATURES  
 source Location/Qualifiers  
 1..50  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G0114A01"  
 /clone\_lib="Mouse E7.5 Embryonic Portion cDNA Library"  
 /sex="unknown"  
 /dev\_stage="7.5dpc Embryo"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1 (Gibco/BRL Life Technology); Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from 6 Embryo. The double-stranded cDNA was synthesized by Gibco's kit with an oligo(dT) primer [NotI primer-adaptor from Gibco/BRL] [5'-pGACTAGTCTAGATCGCGCGCGCTTTT-3'] from 0.5ug of mRNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang and Minoru S. H. Ko."  
 17 a 14 c 6 g 13 t

BASE COUNT 17 a 14 c 6 g 13 t  
 ORIGIN  
 Query Match 57.0%; Score 11.4; DB 117; Length 50;  
 Best Local Similarity 92.3%; Pred. No. 7.3e+04;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ttggtttccaatc 15  
 ||| |||||  
 Db 20 TTGGTTCCAATC 8

RESULT 14  
 AA652673/c 40 bp mRNA EST 25-NOV-1997  
 LOCUS ns69f11.s1 NCI\_CGAP\_Pr2 Homo sapiens cDNA clone IMAGE:118909  
 DEFINITION similar to TR:G1712 G1712 RETROVIRUS RELATED REVERSE TRANSCRIPTASE  
 ;contains L1.t1 L1 repetitive element ;, mRNA sequence.  
 AA652673  
 AA652673.1 GI:2584325  
 EST.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 40)  
 AUTHORS NCI-CGAP  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai, M.D., Michael Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 DNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
 Insert Length: 1233 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1..40  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1188909"  
 /clone\_lib="NCI\_CGAP\_Pr2"  
 /sex="Male"  
 /dev\_stage="45 years old"  
 /lab\_host="DH10B"  
 /note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected preneoplastic cells histologically-determined to be prostatic intraepithelial neoplasia 2 (PIN2) cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 21 a 3 c 9 g 7 t  
 ORIGIN  
 Query Match 56.0%; Score 11.2; DB 10; Length 40;  
 Best Local Similarity 81.2%; Pred. No. 8.9e+04;  
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttggtttccaatcgga 18  
 ||| |||||  
 Db 40 TTCCTTCCAATCNGA 25

RESULT 15  
 AZ822833 43 bp DNA GSS 20-FEB-2001  
 LOCUS 2M0096J23F Mouse 10kb plasmid UGCIW library Mus musculus genomic  
 DEFINITION clone UGCIW2M0096J23 F, DNA sequence.  
 AZ822833  
 AZ822833.1 GI:12992741  
 GSS.  
 KEYWORDS house musculus.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah  
Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0096 row: J column: 23  
Seq primer: CTTGTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 43.

FEATURES

source

1..43  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U0GC2M0096J23"  
/clone\_lib="Mouse 10kb plasmid U0GC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42hv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (g1147321141gb1AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

11 a 8 c 8 g 16 t

Query Match 56.0%; Score 11.2; DB 250; Length 43;  
Best Local Similarity 81.2%; Pred. NO. 9e+04;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 ttggtttccaatcga 18  
| | | | | | | | | |  
Db 21 TGGTTTACACTCGGA 36

Search completed: October 2, 2001, 15:01:04  
Job time: 10833 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:49 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100b-16

Perfect score: 20

Sequence: 1 ctaggaggagctcagtggtgg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 773874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_ba3:\*

4: gb\_in1:\*

5: gb\_in2:\*

6: gb\_in3:\*

7: gb\_om:\*

8: gb\_ov:\*

9: gb\_pat1:\*

10: gb\_pat2:\*

11: gb\_ph:\*

12: gb\_pl1:\*

13: gb\_pl2:\*

14: gb\_pl3:\*

15: gb\_pl4:\*

16: em\_ba1:\*

17: em\_ba2:\*

18: em\_fun:\*

19: em\_htgo\_hum:\*

20: em\_htgo\_inv:\*

21: em\_htgo\_rod:\*

22: em\_htg\_hum1:\*

23: em\_htg\_hum2:\*

24: em\_htg\_hum3:\*

25: em\_htg\_hum4:\*

26: em\_htg\_hum5:\*

27: em\_htg\_hum6:\*

28: em\_htg\_hum7:\*

29: em\_htg\_hum8:\*

30: em\_htg\_inv1:\*

31: em\_htg\_inv2:\*

32: em\_htg\_other:\*

33: em\_htg\_rod:\*

34: em\_humi:\*

35: em\_hum2:\*

36: em\_hum3:\*

37: em\_hum4:\*

38: em\_hum5:\*

39: em\_hum6:\*

40: em\_hum7:\*

41: em\_in:\*

42: em\_om:\*

43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vl:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vil:\*

59: gb\_vil2:\*

60: gb\_htg1:\*

61: gb\_htg2:\*

62: gb\_htg3:\*

63: gb\_htg4:\*

64: gb\_htg5:\*

65: gb\_htg6:\*

66: gb\_htg7:\*

67: gb\_htg8:\*

68: gb\_htg9:\*

69: gb\_htg10:\*

70: gb\_htg11:\*

71: gb\_htg12:\*

72: gb\_htg13:\*

73: gb\_htg14:\*

74: gb\_htg15:\*

75: gb\_htg16:\*

76: gb\_htg17:\*

77: gb\_htg18:\*

78: gb\_htg19:\*

79: gb\_htg20:\*

80: gb\_htg21:\*

81: gb\_htg22:\*

82: gb\_htg23:\*

83: gb\_htg24:\*

84: gb\_htg25:\*

85: gb\_pr1:\*

86: gb\_pr2:\*

87: gb\_pr3:\*

88: gb\_pr4:\*

89: gb\_pr5:\*

90: gb\_pr6:\*

91: gb\_pr7:\*

92: gb\_pr8:\*

93: gb\_pr9:\*

94: gb\_rol:\*

95: gb\_rod:\*

96: gb\_in4:\*

97: gb\_pr10:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13.2	66.0	20	9	AR100356 Sequence
C 2	13.2	66.0	47	9	AR032663 Sequence
C 3	13.2	66.0	47	10	I29403 Sequence 27
C 4	13.2	66.0	47	10	I91077 Sequence 27
C 5	12.8	64.0	30	10	I19158 Sequence 21
6	12.8	64.0	47	9	AR032673 Sequence
7	12.8	64.0	47	10	I29413 Sequence 28
8	12.8	64.0	47	10	I91087 Sequence 28

c 9 12.6 63.0 28 9 AR066591 Sequence  
 c 10 12.6 63.0 28 10 E37410 Recombinant  
 c 11 12.6 63.0 28 10 I06858  
 c 12 12.6 63.0 32 9 AR080756  
 c 13 12.6 63.0 32 10 I28948  
 c 14 12.6 63.0 33 10 I43985  
 c 15 12.6 63.0 42 97 S80775  
 c 16 12.4 62.0 21 10 AX096203  
 c 17 12.4 62.0 30 9 AR100568  
 c 18 12.4 62.0 30 10 I19159  
 c 19 12.2 61.0 30 10 E15705  
 c 20 12.2 61.0 34 9 AX003737  
 c 21 12.2 61.0 34 9 AX023341  
 c 22 12 60.0 16 9 AX061521  
 c 23 12 60.0 29 9 A23243  
 c 24 12 60.0 29 9 A23244  
 c 25 12 60.0 29 9 AR040875  
 c 26 12 60.0 29 9 AR040876  
 c 27 12 60.0 29 9 AR068407  
 c 28 12 60.0 29 9 AR068408  
 c 29 12 60.0 38 9 AR075847  
 c 30 12 60.0 39 9 A51622  
 c 31 12 60.0 40 10 I62865  
 c 32 12 60.0 49 10 E08692  
 c 33 11.8 59.0 19 9 AR021123  
 c 34 11.8 59.0 19 9 AR036267  
 c 35 11.8 59.0 21 9 AX037078  
 c 36 11.8 59.0 22 9 AR101978  
 c 37 11.8 59.0 23 10 I67849  
 c 38 11.8 59.0 26 10 I43986  
 c 39 11.8 59.0 31 9 AX001457  
 c 40 11.8 59.0 33 9 AR063468  
 c 41 11.8 59.0 33 10 I28148  
 c 42 11.8 59.0 40 9 AX027275  
 c 43 11.8 59.0 46 94 MMV66RNA  
 c 44 11.8 59.0 48 9 AR050776  
 c 45 11.8 59.0 48 9 AR050777

## ALIGNMENTS

RESULT 1  
 AR100356  
 LOCUS AR100356 20 bp DNA PAT 14-FEB-2001  
 DEFINITION Sequence 87 from patent US 6080580.  
 ACCESSION AR100356  
 VERSION AR100356.1 GI:12810804  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 20)  
 Baker,B.F., Bennett,C.Frank, Butler,M.M. and Shanahan,W.R. Jr.  
 Antisense oligonucleotide modulation of tumor necrosis  
 factor- $\alpha$ . (TNF- $\alpha$ .) expression  
 JOURNAL Patent: US 6080580-A 87 27-JUN-2000;  
 FEATURES  
 Location/Qualifiers  
 source 1..20  
 BASE COUNT 5 a 3 c 9 g 3 t  
 ORIGIN  
 Query Match 66.0%; Score 13.2; DB 9; Length 20;  
 Best Local Similarity 83.3%; Pred. No. 5.8e+04;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 3 agggaggagctcagtggtg 20  
 ||||| |||||  
 Db 1 AGAGGAGAGTCAGTGTGG 18

RESULT 2  
 AR032663/c  
 LOCUS AR032663 47 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 275 from patent US 5869241.  
 ACCESSION AR032663  
 VERSION AR032663.1 GI:5948268  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 47)  
 Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
 Method of determining DNA sequence preference of a DNA-binding  
 molecule  
 JOURNAL Patent: US 5869241-A 275 09-FEB-1999;  
 FEATURES  
 Location/Qualifiers  
 source 1..47  
 BASE COUNT 12 a 15 c 13 g 7 t  
 ORIGIN  
 Query Match 66.0%; Score 13.2; DB 9; Length 47;  
 Best Local Similarity 83.3%; Pred. No. 4.8e+04;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 2 tagggaggagctcagtggtg 19  
 ||||| ||||| |||||  
 Db 22 TAGGGCAGGCTCTGTTTG 5  
 RESULT 3  
 I29403/c  
 LOCUS I29403 47 bp DNA PAT 06-FEB-1997  
 DEFINITION Sequence 275 from patent US 5578444.  
 ACCESSION I29403  
 VERSION I29403.1 GI:1820194  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 47)  
 Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
 Sequence-directed DNA-binding molecules compositions and methods  
 JOURNAL Patent: US 5578444-A 275 26-NOV-1996;  
 FEATURES  
 Location/Qualifiers  
 source 1..47  
 BASE COUNT 12 a 15 c 13 g 7 t  
 ORIGIN  
 Query Match 66.0%; Score 13.2; DB 10; Length 47;  
 Best Local Similarity 83.3%; Pred. No. 4.8e+04;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 2 tagggaggagctcagtggtg 19  
 ||||| ||||| |||||  
 Db 22 TAGGGCAGGCTCTGTTTG 5  
 RESULT 4  
 I91077/c  
 LOCUS I91077 47 bp DNA PAT 01-DEC-1998  
 DEFINITION Sequence 275 from patent US 5726014.  
 ACCESSION I91077  
 VERSION I91077.1 GI:3935547  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 47)  
 Edwards,C.A., Cantor,C.R., Andrews,B.M. and Turin,L.M.

TITLE Screening assay for the detection of DNA-binding molecules  
JOURNAL Patent: US 5726014-A 275 10-MAR-1998;  
FEATURES Location/Qualifiers

source  
1. .47  
/organism="unknown"

BASE COUNT 12 a 15 c 13 g 7 t  
ORIGIN

Query Match 66.0%; Score 13.2; DB 10; Length 47;  
Best Local Similarity 83.3%; Pred. No. 4.8e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tagggaggctcagtggtg 19  
||||| ||||| |||||  
Db 22 TAGGGCAGGCTCTGTTG 5

RESULT 5  
I19158/c  
LOCUS 119158 30 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 21 from patent US 5502176.  
ACCESSION I19158  
VERSION I19158.1 GI:1599513  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 30)  
AUTHORS Tenen,D.G.; Pahl,H.L. and Burn,T.C.  
TITLE Myeloid cell specific promoter  
JOURNAL Patent: US 5502176-A 21 26-MAR-1996;  
FEATURES Location/Qualifiers  
source  
1. .30  
/organism="unknown"

BASE COUNT 2 a 13 c 5 g 10 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 10; Length 30;  
Best Local Similarity 87.5%; Pred. No. 8.4e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ggggaggctcagtggtg 19  
||||| ||||| |||||  
Db 18 GCGGAGGCTCAGTGAG 3

RESULT 6  
AR032673  
LOCUS AR032673 47 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 285 from patent US 5869241.  
ACCESSION AR032673  
VERSION AR032673.1 GI:5948278  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 47)  
AUTHORS Edwards,C.A.; Cantor,C.R.; Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule  
JOURNAL Patent: US 5869241-A 285 09-FEB-1999;  
FEATURES Location/Qualifiers  
source  
1. .47  
/organism="unknown"

BASE COUNT 14 a 15 c 12 g 6 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 9; Length 47;  
Best Local Similarity 87.5%; Pred. No. 7.6e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 aggggaggctcagtggt 18  
||||| ||||| |||||  
Db 8 AGGGAAGGCTCAGTAT 23

RESULT 7  
I29413  
LOCUS I29413 47 bp DNA PAT 06-FEB-1997  
DEFINITION Sequence 285 from patent US 5578444.  
ACCESSION I29413  
VERSION I29413.1 GI:1820204  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 47)  
AUTHORS Edwards,C.A.; Cantor,C.R.; Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Sequence-directed DNA-binding molecules compositions and methods  
JOURNAL Patent: US 5578444-A 285 26-NOV-1996;  
FEATURES Location/Qualifiers  
source  
1. .47  
/organism="unknown"

BASE COUNT 14 a 15 c 12 g 6 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 10; Length 47;  
Best Local Similarity 87.5%; Pred. No. 7.6e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 aggggaggctcagtggt 18  
||||| ||||| |||||  
Db 8 AGGGAAGGCTCAGTAT 23

RESULT 8  
I91087  
LOCUS I91087 47 bp DNA PAT 01-DEC-1998  
DEFINITION Sequence 285 from patent US 5726014.  
ACCESSION I91087  
VERSION I91087.1 GI:3935557  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 47)  
AUTHORS Edwards,C.A.; Cantor,C.R.; Andrews,B.M. and Turin,L.M.  
TITLE Screening assay for the detection of DNA-binding molecules  
JOURNAL Patent: US 5726014-A 285 10-MAR-1998;  
FEATURES Location/Qualifiers  
source  
1. .47  
/organism="unknown"

BASE COUNT 14 a 15 c 12 g 6 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 10; Length 47;  
Best Local Similarity 87.5%; Pred. No. 7.6e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 aggggaggctcagtggt 18  
||||| ||||| |||||  
Db 8 AGGGAAGGCTCAGTAT 23

RESULT 9  
AR066591/c  
LOCUS AR066591 28 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 2 from patent US 5851529.  
ACCESSION AR066591  
VERSION AR066591.1 GI:5997813  
KEYWORDS

SOURCE ORGANISM Unknown.  
Unknown.  
REFERENCE 1 (bases 1 to 28)  
AUTHORS Guber,H.E., Jolly,D.J., Respass,J.G. and Laikind,P.K.  
TITLE Recombinant retroviruses  
JOURNAL Patent: US 5851529-A 2 22-DEC-1998;  
FEATURES Location/Qualifiers  
source 1..28  
BASE COUNT 4 a 13 c 3 g 8 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 28;  
Best Local Similarity 78.9%; Pred. No. 1.le+05;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtggtg 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 22 TGGGGAGGCTAACTGAGG 4

RESULT 10  
E37410/c 28 bp DNA PAT 07-FEB-2001  
LOCUS Recombinant retrovirus with defective replication.  
DEFINITION E37410  
ACCESSION E37410  
VERSION 1 GI:130233106  
KEYWORDS JP 1999262397-A/2.  
SOURCE unidentified.  
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 28)  
AUTHORS Halley,E.G.D.D., Jolly,J.G.R.R. and K.R.  
TITLE Recombinant retrovirus with defective replication  
JOURNAL Patent: JP 1999262397-A 2 28-SEP-1999;  
COMMENT CHIRON CORP  
OS Artificial Sequence  
PN JP 1999262397-A/2  
PD 28-SEP-1999  
PF 24-DEC-1998 JP 1998368260  
PR 21-MAR-1988 US 170.515  
PI HALLEY E GRABAR, DOUGLAS J JOLLY, JAMES G RESPASU, PAUL K REIKAINEN  
PC C12N15/09, A01K67/027, A61K31/00, A61K35/12, A61K35/76,  
PC A61K48/00,  
PC C12N5/10, C12N7/00, C12P21/00, C12R1/91, C12N15/00,  
CC C12N5/00

FEATURES  
FH Key Location/Qualifiers  
FT source 1..28  
FT /organism="Artificial Sequence".  
source Location/Qualifiers  
1..28  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 4 a 13 c 3 g 8 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 28;  
Best Local Similarity 78.9%; Pred. No. 1.le+05;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtggtg 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 22 TGGGGAGGCTAACTGAGG 4

RESULT 11  
I06858/c 28 bp PAT 02-DEC-1994  
LOCUS Sequence 2 from Patent EP 0334301.  
DEFINITION

ACCESSION I06858  
VERSION I06858.1 GI:589980  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 28)  
AUTHORS Gruber,H.E., Jolly,D.J., Respass,J.G. and Laikind,P.K.  
TITLE Recombinant retroviruses  
JOURNAL Patent: EP 0334301-A1 2 27-SEP-1989;  
FEATURES Location/Qualifiers  
source 1..28  
BASE COUNT 4 a 13 c 3 g 8 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 28;  
Best Local Similarity 78.9%; Pred. No. 1.le+05;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtggtg 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 22 TGGGGAGGCTAACTGAGG 4

RESULT 12  
AR080756/c 32 bp DNA PAT 31-AUG-2000  
LOCUS AR080756  
DEFINITION Sequence 2 from patent US 5968828.  
ACCESSION AR080756  
VERSION AR080756.1 GI:10007486  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Pehu,E., Pehu,T., Maki-Valkama,T., Valkonen,J., Koivu,K. and Lehto,K.  
TITLE Virus-resistant transgenic plants comprising cells transformed with a polynucleotide encoding a potyvirus P1 protein or P1 protein fragment  
JOURNAL Patent: US 5968828-A 2 19-OCT-1999;  
FEATURES Location/Qualifiers  
source 1..32  
/organism="unknown"  
BASE COUNT 8 a 8 c 9 g 7 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 32;  
Best Local Similarity 78.9%; Pred. No. 1.le+05;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctggggaggctcagtggtg 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 29 CTAGGTGAGCTCATTTAG 11

RESULT 13  
I28948/c 32 bp DNA PAT 06-FEB-1997  
LOCUS I28948  
DEFINITION Sequence 2 from patent US 5576202.  
ACCESSION I28948  
VERSION I28948.1 GI:1819739  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Pehu,E., Pehu,T., Maki-Valkama,T., Valkonen,J., Koivu,K. and Lehto,K.  
TITLE Virus-resistant transgenic plants

JOURNAL Patent: US 5576202-A 2 19-NOV-1996;

FEATURES Location/Qualifiers

source 1. .32

BASE COUNT 8 a 8 c 9 g 7 t

ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 32;

Best Local Similarity 78.9%; Pred. No. 1e+05; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctaggggaggctcagtggtg 19

||||| ||| ||||| ||

Db 29 CTAGGTGAGACTCATTTGAG 11

RESULT 14

I43985/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 33;

Best Local Similarity 78.9%; Pred. No. 1e+05; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggggaggctcagtggtg 20

||||| ||| ||||| ||

Db 21 TATGGGAGCTCAGCTAGG 3

RESULT 15

S80775

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

FEATURES

source

BASE COUNT

ORIGIN

/db\_xref="taxon:9606"

1. .42

/partial

/gene="gamma delta T cell antigen receptor delta-chain"

10 a 9 c 17 g 6 t

BASE COUNT

ORIGIN

Query Match 63.0%; Score 12.6; DB 97; Length 42;

Best Local Similarity 78.9%; Pred. No. 9.9e+04;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctaggggaggctcagtggtg 19

||||| ||| ||||| ||

Db 14 CTGGGGGAGCCCCAGTGAG 32

Search completed: October 2, 2001, 15:56:49

Job time: 14173 sec

/organism="Homo sapiens"





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:04 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-16  
Perfect score: 20  
Sequence: 1 ctaggggaggtcagtggtgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	Count				
1	13.2	66.0	50	107	AU107335	AU107335	AU107335
2	13.2	66.0	50	107	AU107336	AU107336	AU107336
3	12.6	63.0	33	250	A2826713	A2826713	A2826713
C 4	12.6	63.0	46	102	A1809473	A1809473	A1809473
5	12.6	63.0	42	245	A2498822	A2498822	A2498822
6	12.6	60.0	48	250	A2833039	A2833039	A2833039
C 7	11.8	59.0	21	247	A2660221	A2660221	A2660221
C 8	11.8	59.0	50	107	AU104075	AU104075	AU104075
C 9	11.8	59.0	50	107	AU105603	AU105603	AU105603
C 10	11.8	59.0	50	107	AU105604	AU105604	AU105604
C 11	11.8	59.0	50	107	AU105605	AU105605	AU105605
C 12	11.8	59.0	50	107	AU105606	AU105606	AU105606
C 13	11.6	58.0	20	243	A2417155	A2417155	A2417155
C 14	11.6	58.0	26	258	TA91803P	TA91803P	TA91803P
C 15	11.6	58.0	40	189	T95881	T95881	T95881
C 16	11.6	58.0	42	167	BE386700	BE386700	BE386700
17	11.6	58.0	43	13	A2323260	A2323260	A2323260
C 18	11.6	58.0	46	19	A1360975	A1360975	A1360975
C 19	11.6	58.0	47	139	BE738401	BE738401	BE738401
C 20	11.6	58.0	48	139	BE738353	BE738353	BE738353
C 21	11.6	58.0	48	139	BE738724	BE738724	BE738724
C 22	11.6	58.0	50	107	AU106034	AU106034	AU106034
C 23	11.4	57.0	31	14	A1000255	A1000255	A1000255
C 24	11.4	57.0	36	244	A2484337	A2484337	A2484337
C 25	11.4	57.0	40	246	AZ597058	AZ597058	AZ597058
C 26	11.4	57.0	44	250	AZ818897	AZ818897	AZ818897
C 27	11.2	56.0	24	250	AZ844396	AZ844396	AZ844396
C 28	11.2	56.0	30	250	AZ853328	AZ853328	AZ853328
C 29	11.2	56.0	37	242	TA116D11P	TA116D11P	TA116D11P
C 30	11.2	56.0	41	258	TA116D11P	TA116D11P	TA116D11P
C 31	11.2	56.0	42	246	AZ592675	AZ592675	AZ592675
C 32	11.2	56.0	43	4A	A238745	A238745	A238745
C 33	11.2	56.0	46	159	N24850	N24850	N24850
C 34	11	55.0	22	246	A2623308	A2623308	A2623308
C 35	11	55.0	35	258	TA177A10Q	TA177A10Q	TA177A10Q
C 36	11	55.0	37	13	A933928	A933928	A933928
C 37	11	55.0	39	247	AZ659784	AZ659784	AZ659784
C 38	11	55.0	41	244	AZ470576	AZ470576	AZ470576
C 39	11	55.0	42	243	AZ440678	AZ440678	AZ440678
C 40	11	55.0	43	247	AZ660216	AZ660216	AZ660216
C 41	11	55.0	44	158	H92874	H92874	H92874
C 42	11	55.0	49	241	A3133961	A3133961	A3133961
C 43	11	55.0	49	258	TA307F09Q	TA307F09Q	TA307F09Q
C 44	11	55.0	50	107	AU106627	AU106627	AU106627
C 45	10.8	54.0	19	249	AZ775541	AZ775541	AZ775541

## ALIGNMENTS

RESULT	1
LOCUS	AU107335
DEFINITION	50 bp mRNA EST 05-APR-2001
	AU107335 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
ACCESSION	KAT02360, mRNA sequence.
VERSION	AU107335
KEYWORDS	AU107335.1 GI:13556856
SOURCE	EST.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 50)
AUTHORS	Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata H., Ota,T., Isozaki,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo K., Suyama,A. and Sugano,S.
TITLE	Fine Structural analysis of transcription start sites of human

mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
Unpublished (2001)  
Contact: Yutaka Suzuki

Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yuzuki@ims.u-tokyo.ac.jp  
Suzuki, Y., Yoshimoto-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* 200 (1-2), 149-156 (1997).

```

FEATURES
  source
    Location/Qualifiers
      1..50
        organism="Homo sapiens"
        db_xref="taxon:9606"
        clone="KAT02360"
        clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT
  6 a 12 c 25 g 7 t

```

Query Match	66.0%	Score 13.2;	DB 107;	Length 50;
Best Local Similarity	83.3%	Pred. No. 3.7e+04;		
Matches 15; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

Qy	1	ctaggggagggctcagtg	18
Db	31	CTAGGGCGGCTCGGAGT	48

RESULT	2
AU107336	
LOCUS	AU107336 50 bp mRNA EST 05-APR-2001
DEFINITION	AU107336 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone LN11142, mRNA sequence.
ACCESSION	AU107336
VERSION	AU107336.1
KEYWORDS	GI:13556857
SOURCE	EST.
	human.

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 50)  
 REFERENCE  
 SUZUKI, Y., TSUNODA, T., TAIRA, H., MIZUSHIMA-SUGANO, J., SESE, J., HATA  
 K., OTA, T., ISOGAI, T., TANAKA, T., NAKAMURA, Y., MORISHITA, S., OKUBO  
 H., SUYAMA, A. AND SUGANO, S.

**TITLE** Fine Structural analysis of transcription start sites of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries

**JOURNAL** Unpublished (2001)

**COMMENT** Contact: Yutaka Suzuki

Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yszukui@ims.u-tokyo.ac.jp  
Suzuki, Y., Yoshimoto-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* 200 (1-2): 149-156 (1997).

```

FEATURES             source
    1..50
        Location/Qualifiers
            a 3'-end-enriched cDNA library. Gene 200 (1-2), 145-150 (1997).
            Location/Qualifiers
                6..112..35..94
                    clone_lib="Sugano Homo sapiens cDNA library"
                    db_xref="taxon:9606"
                    clone="LNG11142"
                    /organism="Homo sapiens"

```

Query Match	66.0%	Score 13.2	DB 107	Length 50
Best Local Similarity	83.3%	Pred. No. 3.7e+04		
Matches 15	Conservative	0	Mismatches 3	Indels 0
				Caps 0

Qy 1 ctaggggaggctcagtg 18  
|||  
32 CTAGGGCGGCTCGAGT 49  
Db

```

Db 11 CTAGGTGCCGCGAGTGTG 29

RESULT 3
LOCUS AZ826713 33 bp DNA GSS 20-FEB-2001
DEFINITION 2M0102H12R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGC2M0102H12 R, DNA sequence.
ACCESSION AZ826713
VERSION AZ826713.1 GI:12996621
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 33)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0102 row: H column: 12
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 33.
Location/Qualifiers
1. .33
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0102H12"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gbIAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
6 a 8 c 13 g 6 t

BASE COUNT
ORIGIN

Query Match 63.0%; Score 12.6; DB 250; Length 33;
Best Local Similarity 78.9%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctaggaggagctcagtgctg 19
||||| | || |||||

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred. No. 7.1e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtgctg 20
||||| | ||||| |||||
Db 46 TAGGGGTGTTCCAGGGTGG 28

RESULT 5
LOCUS AZ498822 42 bp DNA GSS 05-OCT-2000
DEFINITION 1M0336117F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0336117 F, DNA sequence.
ACCESSION AZ498822
VERSION AZ498822.1 GI:10677034
KEYWORDS GSS.

BASE COUNT 14 a 25 c 1 g 6 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 102; Length 46;
Best Local Similarity 78.9%; Pred.
```

**SOURCE** house mouse.  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 42)  
**AUTHORS** Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0336 row: 1 column: 17  
 Seq primer: CGTGTAAACGAGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 42.  
**FEATURES**  
 Location/Qualifiers  
 1..42  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0336117"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
**BASE COUNT** 8 a 2 c 25 g 7 t  
**ORIGIN**  
 Query Match 60.0%; Score 12; DB 245; Length 42;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+05;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 ctaggggaggctcagtgtg 20  
 |||||  
 Db 20 CTAGGGAGAGCTGGGGGGG 39  
**RESULT** 6  
 AZ833039 48 bp DNA GSS 20-FEB-2001  
 LOCUS 2M0113P15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC2M0113P15 R, DNA sequence.  
 ACCESSION AZ833039  
 VERSION AZ833039.1 GI:13002947

**KEYWORDS** - GSS.  
**SOURCE** house mouse.  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 48)  
**AUTHORS** Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0113 row: P column: 15  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 48.  
**FEATURES**  
 Location/Qualifiers  
 1..48  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0113P15"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWB42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
**BASE COUNT** 16 a 8 c 17 g 7 t  
**ORIGIN**  
 Query Match 60.0%; Score 12; DB 250; Length 48;  
 Best Local Similarity 75.0%; Pred. No. 1.4e+05;  
 Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 ctaggggaggctcagtgtg 20  
 |||||  
 Db 7 CTGGGGGAGGGGCGTGGGG 26  
**RESULT** 7  
 AZ660221/c 21 bp DNA GSS 14-DEC-2000  
 LOCUS IM0538M09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC1M0538M09 F, DNA sequence.  
 ACCESSION AZ660221

```

VERSION      AZ660221.1  GI:11797367
KEYWORDS     GSS.
SOURCE       Mus musculus
ORGANISM     Mus musculus
REFERENCE    1 (bases 1 to 21)
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
             Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
             ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
             and Wright,D.,Weiss,R.
TITLE        Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 309, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: dunn@genetics.utah.edu
             Insert Length: 10000 Std Error: 0.00
             Plate: 0538 row: M column: 09
             Seq primer: CGTTGTAACGACGGCCAGT
             Class: plasmid ends
             High quality sequence stop: 21.
FEATURES     Location/Qualifiers
             1..21
             /organism="Mus musculus"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="UUGC1M0538M09"
             /clone.lib="Mouse 10kb plasmid UUGC1M library"
             /sex="Male"
             /lab.host="E. Coli strain XL10-Gold, T1-resistant, F-"
             /note="Vector: PWD42hv; Purified genomic DNA from M.
             musculus C57BL/6J (male) was obtained from the Jackson
             Laboratory Mouse DNA Resource
             (http://www.jax.org/resources/documents/dnares/). The DNA
             was hydrodynamically sheared by repeated passage through a
             0.005 inch orifice at constant velocity. The sheared DNA
             was blunt end-repaired with T4 DNA polymerase and T4
             polynucleotide kinase. Adaptor oligonucleotides were
             ligated to the blunt ends in high molar excess. The
             adapted DNA was purified and size-selected for a 9.5 to
             10.5 kb range using preparative agarose gel
             electrophoresis. Vector DNA was prepared from a derivative
             of PWD42 (g114732114[gblAF129072.1], a copy-number
             inducible derivative of plasmid R1. The vector was ligated
             with adaptors complementary to the insert adaptors and
             purified. The sheared, adapted mouse DNA was annealed to
             adapted vector DNA, and transformed into
             chemically-competent E. coli XL10-Gold (Stratagene) cells
             and selected for ampicillin resistance."
BASE COUNT   7 a 11 c 1 g 2 t
ORIGIN
Query Match      59.0%; Score 11.8; DB 247; Length 21;
Best Local Similarity 86.7%; Pred. No. 1.6e+05;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 gggagggcagtggtg 19
    ||| ||| ||| ||| |||
Db 16 GGGTGGTCAGTGTG 2

RESULT 8
AUI04075 50 bp mRNA EST 05-APR-2001
LOCUS
DEFINITION AUI04075 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP14513, mRNA sequence.
ACCESSION AUI04075.1 GI:13553596
VERSION AUI04075.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 50)
AUTHORS H., Ota,T., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
        ,K., Suyama,A. and Sugano,S.
        Fine Structural analysis of transcription start sites of human
        mRNAs using full-length enriched and 5'-end enriched cDNA libraries
        Unpublished (2001)
        Contact: Yutaka Suzuki
        Department of Virology
        Institute of Medical Science, University of Tokyo
        4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
        Email: ysuzuki@ims.u-tokyo.ac.jp
        Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
        ,S. Construction and characterization of a full length-enriched and
        a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES     Location/Qualifiers
             1..50
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="HEP14513"
             /clone.lib="Sugano Homo sapiens cDNA library"
BASE COUNT   8 a 11 c 20 g 11 t
ORIGIN
Query Match      59.0%; Score 11.8; DB 107; Length 50;
Best Local Similarity 86.7%; Pred. No. 1.7e+05;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agggagggcagtggtg 17
    ||| ||| ||| ||| |||
Db 17 AGGGGAGGCGCGGTG 31

RESULT 9
AUI05603 50 bp mRNA EST 05-APR-2001
LOCUS
DEFINITION AUI05603 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HSI00510, mRNA sequence.
ACCESSION AUI05603.1 GI:13555124
VERSION AUI05603.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 50)
AUTHORS H., Ota,T., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
        ,K., Suyama,A. and Sugano,S.
        Fine Structural analysis of transcription start sites of human
        mRNAs using full-length enriched and 5'-end enriched cDNA libraries
        Unpublished (2001)
        Contact: Yutaka Suzuki
        Department of Virology
        Institute of Medical Science, University of Tokyo
        4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
        Email: ysuzuki@ims.u-tokyo.ac.jp
        Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
        ,S. Construction and characterization of a full length-enriched and
        a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES     Location/Qualifiers
             1..50
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="HSI00510"

```

BASE COUNT 9 a 18 c 11 g 12 t  
ORIGIN /clone\_lib="Sugano Homo sapiens cDNA library"

Query Match 59.0%; Score 11.8; DB 107; Length 50;  
Best Local Similarity 86.7%; Pred. No. 1.7e+05;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 gggaggctcagtg 19  
||||| ||||| |||||

Db 41 GGGAGCCTCATGTG 27

RESULT 10

AU105604/c

LOCUS AU105604 50 bp mRNA EST 05-APR-2001

DEFINITION AU105604 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone

HS101710, mRNA sequence.

ACCESSION AU105604

VERSION AU105604.1 GI:13555125

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 50)

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata

,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo

,K., Suyama,A. and Sugano,S.

Fine structural analysis of transcription start sites of human

mRNAs using full-length enriched and 5'-end enriched cDNA libraries

Unpublished (2001)

Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: yszukie@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano

,S. Construction and characterization of a full length-enriched and

a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1. .50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HS101710"

/clone\_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 10 a 18 c 8 g 14 t

ORIGIN

Query Match 59.0%; Score 11.8; DB 107; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.7e+05;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 gggaggctcagtg 19

||||| ||||| |||||

Db 44 GGGAGCCTCATGTG 30

RESULT 11

AU105605/c

LOCUS AU105605 50 bp mRNA EST 05-APR-2001

DEFINITION AU105605 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone

HS105558, mRNA sequence.

ACCESSION AU105605

VERSION AU105605.1 GI:13555126

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 50)

AUTHORS

Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata

,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo

,K., Suyama,A. and Sugano,S.

Fine structural analysis of transcription start sites of human

mRNAs using full-length enriched and 5'-end enriched cDNA libraries

Unpublished (2001)

Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: yszukie@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano

,S. Construction and characterization of a full length-enriched and

a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

Location/Qualifiers

1. .50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HS105558"

/clone\_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 10 a 18 c 9 g 13 t

ORIGIN

Query Match 59.0%; Score 11.8; DB 107; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.7e+05;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 gggaggctcagtg 19

||||| ||||| |||||

Db 42 GGGAGCCTCATGTG 28

RESULT 12

AU105606/c

LOCUS AU105606 50 bp mRNA EST 05-APR-2001

DEFINITION AU105606 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone

HS106872, mRNA sequence.

ACCESSION AU105606

VERSION AU105606.1 GI:13555127

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 50)

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata

,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo

,K., Suyama,A. and Sugano,S.

Fine structural analysis of transcription start sites of human

mRNAs using full-length enriched and 5'-end enriched cDNA libraries

Unpublished (2001)

Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: yszukie@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano

,S. Construction and characterization of a full length-enriched and

a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

Location/Qualifiers

1. .50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HS106872"

/clone\_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 10 a 18 c 10 g 12 t

ORIGIN

Query Match 59.0%; Score 11.8; DB 107; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.7e+05;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 5 gggaggctcagtg 19
    ||||| ||||| |||||
Db 40 GGGAGCTCATTTGTG 26

RESULT 13
LOCUS      A2417155/c      20 bp      DNA      GSS      03-OCT-2000
DEFINITION LM0192P07R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
            clone UUGCLM0192P07 R, DNA sequence.
ACCESSION  A2417155
VERSION    A2417155.1 GI:10541168
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 20)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0192 row: P column: 07
Seq primer: CACAGAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 20.

FEATURES             source
1..20
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0192P07"
/clone.lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g1147321141gb1AF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      8 a      7 c      2 g      3 t
ORIGIN

Query Match      58.0%; Score 11.6; DB 243; Length 20;
Best Local Similarity 77.8%; Pred. No. 2e+05;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tagggaggctcagtg 19
    ||||| ||||| |||||
Db 20 TAGGGGTAGCTCTTTGTG 3

RESULT 14
LOCUS      TA91B03P      26 bp      DNA      GSS      13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 91b03, forward sequence,
            genomic survey sequence.
ACCESSION  AL459576
VERSION    AL459576.1 GI:11861682
KEYWORDS   GSS.
SOURCE     Trypanosoma brucei.
ORGANISM   Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
1 (bases 1 to 26)
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Meiville,S.E., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhesanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES             source
1..26
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="91b03"
BASE COUNT      6 a      8 c      8 g      4 t
ORIGIN

Query Match      58.0%; Score 11.6; DB 258; Length 26;
Best Local Similarity 77.8%; Pred. No. 2e+05;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctaggggaggctcagtg 18
    | |||| | |||| | ||||
Db 2 CAAGGGCAGGATCGCTGT 19

RESULT 15
LOCUS      T95881/c      40 bp      mRNA      EST      27-MAR-1995
DEFINITION ye47b02.sl Soares fetal liver spleen INFLS Homo sapiens cDNA clone
            IMAGE:120843 3' similar to gb|M87923|HUMALCE12 Human carcinoma
            cell-derived Alu RNA transcript, (rRNA); gb:M57627 INTERLEUKIN-10
            PRECURSOR (HUMAN); mRNA sequence.
ACCESSION  T95881
VERSION    T95881.1 GI:734505
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match      58.0%; Score 11.6; DB 243; Length 20;
Best Local Similarity 77.8%; Pred. No. 2e+05;

```



1 (bases 1 to 40)  
Hillier,T., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman  
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,  
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston  
R., Williamson,A., Wohldmann,P. and Willson,R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilison RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 2754  
High quality sequence starts: 1 High quality sequence stops: 1  
Source: IMAGE Consortium, LLNL This clone is available royalty-free  
through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov)  
for further information. Trace considered overall poor quality  
Insert Length: 2754 Std Error: 0.00  
Seq Primer: -21ml13

```

1. .40
/organism="Homo sapiens"
/db_xref="GBB:473388"
/db_xref="taxon:9606"
/clone="IMAGE:120843"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I...oligo(30) primer
[5', AACGCGAGAAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
8 a 14 c 10 g 2 t 2 others

```

```

Query Match      58.0%; Score 11.6; DB 189; Length 40;
Best Local Similarity 70.0%; Pred. No. 2.1e+05;
Matches 14; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ctaggggaggctcagtggtg 20
   || ||||| |||||
DB 21 CTTNNGGAGGCTGAGGGGG 2

```

Search completed: October 2, 2001, 15:01:06  
Job time: 10835 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:49 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-17  
Perfect score: 20  
Sequence: 1 attctcgtctgtgtggaa 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rol2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	15.2	76.0	40 9	AR012105	AR012105 Sequence
c 2	15.2	76.0	40 9	AR014544	AR014544 Sequence
c 3	14.2	71.0	20 9	AR059098	AR059098 Sequence
4	13.8	69.0	46 97	HUMTCIG1A	M96133 Human hybri
5	13.6	68.0	32 9	A73580	A73580 Sequence 4
c 6	13.4	67.0	18 9	AR067419	AR067419 Sequence
c 7	13.4	67.0	26 9	AR070795	AR070795 Sequence
c 8	13.4	67.0	39 93	HSCD85706	Z50819 H.sapiens m

c 9 13 65.0 23 9 AX044085 Sequence  
 c 10 13 65.0 23 9 AX044138 Sequence  
 c 11 13 65.0 23 9 AX044178 Sequence  
 c 12 13 65.0 48 92 HSA224214  
 c 13 12.6 63.0 20 9 AR103323  
 c 14 12.6 63.0 30 10 I25076  
 c 15 12.6 63.0 30 10 I71129  
 c 16 12.6 63.0 37 9 A82991  
 c 17 12.6 63.0 37 9 A82993  
 c 18 12.4 62.0 21 9 AX027268  
 c 19 12.2 61.0 20 10 I73108  
 c 20 12.2 61.0 24 9 AR018080  
 c 21 12.2 61.0 24 9 AR051920  
 c 22 12.2 61.0 29 9 AR049382  
 c 23 12.2 61.0 29 9 AR095543  
 c 24 12.2 61.0 30 9 A83263  
 c 25 12.2 61.0 36 93 HSAJ5143  
 c 26 12.2 61.0 38 9 AR108637  
 c 27 12.2 61.0 42 9 AR058955  
 c 28 12.2 61.0 42 9 AR105229  
 c 29 12.2 61.0 49 10 AX098901  
 c 30 12.2 61.0 49 10 AX099321  
 c 31 12 60.0 20 10 E26600  
 c 32 12 60.0 26 10 E04774  
 c 33 12 60.0 27 9 AR090975  
 c 34 12 60.0 28 9 AR050042  
 c 35 12 60.0 28 9 AR064551  
 c 36 12 60.0 31 9 AR013956  
 c 37 12 60.0 31 10 I23438  
 c 38 12 60.0 31 10 I25985  
 c 39 12 60.0 33 9 A68960  
 c 40 12 60.0 45 9 AR065636  
 c 41 12 60.0 45 9 AR068600  
 c 42 12 60.0 45 9 AR095952  
 c 43 12 60.0 45 10 I23529  
 c 44 11.8 59.0 18 9 AR078900  
 c 45 11.8 59.0 23 9 AR106304

## ALIGNMENTS

RESULT 1  
 AR012105/c  
 LOCUS AR012105 40 bp DNA PAT 04-DEC-1998  
 DEFINITION Sequence 52 from patent US 5763209.  
 ACCESSION AR012105  
 VERSION AR012105.1 GI:3970095  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 40)  
 AUTHORS Sukhtame,V.P.  
 TITLE Methods and materials relating to the functional domains of DNA binding proteins  
 JOURNAL Patent: US 5763209-A 52 09-JUN-1998;  
 FEATURES Location/Qualifiers  
 source 1..40  
 BASE COUNT 9 a 14 c 13 g 4 t  
 ORIGIN

Query Match 76.0%; Score 15.2; DB 9; Length 40;  
 Best Local Similarity 85.0%; Pred. No. 4.6e+04;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 attctcgtcgtcgtggtgaa 20  
 ||||| ||||| |||||

Db 21 ATTCTCGGTGCTGCCGGA 2

RESULT 2  
 AR014544/c  
 LOCUS AR014544 40 bp DNA PAT 05-DEC-1998  
 DEFINITION Sequence 52 from patent US 5773583.  
 ACCESSION AR014544  
 VERSION AR014544.1 GI:3971998  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 40)  
 AUTHORS Sukhtame,V.P.  
 TITLE Methods and materials relating to the functional domains of DNA binding proteins  
 JOURNAL Patent: US 5773583-A 52 30-JUN-1998;  
 FEATURES Location/Qualifiers  
 source 1..40  
 BASE COUNT 9 a 14 c 13 g 4 t  
 ORIGIN

Query Match 76.0%; Score 15.2; DB 9; Length 40;  
 Best Local Similarity 85.0%; Pred. No. 4.6e+04;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 attctcgtcgtcgtggtgaa 20  
 ||||| ||||| ||||| |||||

Db 21 ATTCTCGGTGCTGCCGGA 2

RESULT 3  
 AR059098/c  
 LOCUS AR059098 20 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 16 from patent US 5837854.  
 ACCESSION AR059098  
 VERSION AR059098.1 GI:5984675  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Mulder,C.  
 TITLE Oligonucleotides with anti-Epstein-Barr virus activity  
 JOURNAL Patent: US 5837854-A 16 17-NOV-1998;  
 FEATURES Location/Qualifiers  
 source 1..20  
 BASE COUNT 5 a 7 c 4 g 4 t  
 ORIGIN

Query Match 71.0%; Score 14.2; DB 9; Length 20;  
 Best Local Similarity 84.2%; Pred. No. 1.5e+05;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttctcgtcgtcgtggtgaa 20  
 ||||| ||||| ||||| |||||

Db 20 TTGCTCGCAGCTGGAGAA 2

RESULT 4  
 HUMTCIGC1A  
 LOCUS HUMTCIGC1A 46 bp DNA PRI 13-JAN-1995  
 DEFINITION Human hybrid T cell receptor/immunoglobulin protein, clone 1 V-D-J segment.  
 ACCESSION M96133  
 VERSION M96133.1 GI:339188  
 KEYWORDS T-cell receptor/immunoglobulin protein.  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.



```

RESULT 8
HSCD85706/c 39 bp mRNA PRI 15-MAR-1996
LOCUS H.sapiens mRNA for T cell receptor (ID CD85706).
DEFINITION Z50819
ACCESSION Z50819.1 GI:1004241
VERSION constant region; joining region; T cell receptor; variable region.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 39)
AUTHORS Wang, E.C., Moss, P.A., Frodsham, P., Lehner, P.J., Bell, J.I. and
Borysiewicz, L.K.
TITLE Oligoclonal CD8+, CD57+ T cells in normal, healthy individuals and
their responses to HCMV
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39)
AUTHORS Wang, E.C.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1995) Eddie CY Wang, Medicine, University of
Wales of College of Medicine, Heath Park, Cardiff, South
Glamorgan, CF4 4XN, Wales UK
REFERENCE 3 (bases 1 to 39)
AUTHORS Wang, E.C., Moss, P.A., Frodsham, P., Lehner, P.J., Bell, J.I. and
Borysiewicz, L.K.
TITLE CD8highCD57+ T lymphocytes in normal, healthy individuals are
oligoclonal and respond to human cytomegalovirus
JOURNAL J. Immunol. 155 (10), 5046-5056 (1995)
MEDLINE 96062329
FEATURES
source Location/Qualifiers
1..39
/organism="Homo sapiens"
/isolate="EW"
/db_xref="taxon:9606"
1..12
/note="end T cell receptor V beta 24"
13..24
/note="CDR3"
25..39
/note="beginning J beta 2.5"
BASE COUNT 9 a 11 c 15 g 4 t
ORIGIN
Query Match 67.0%; Score 13.4; DB 93; Length 39;
Best Local Similarity 93.3%; Pred. No. 2.7e+05;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 cctcgctgctggtg 18
|||||
Db 19 CCTCTCTGCTGCTGG 5

RESULT 9
AX044085/c 23 bp DNA PAT 24-NOV-2000
LOCUS AX044085
DEFINITION Sequence 40 from Patent WO0066748.
ACCESSION AX044085
VERSION AX044085.1 GI:11342963
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 23)
AUTHORS Hawkes, T.R., Warner, S.A., Andrews, C.J., Bachoo, S. and
Pickerill, A.P.
TITLE Herbicide resistant plants
JOURNAL Patent: WO 0066748-A 40 09-NOV-2000;
ZENECA LIMITED (GB)
FEATURES
source Location/Qualifiers
1..23
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer"
BASE COUNT 5 a 7 c 5 g 6 t
ORIGIN
Query Match 65.0%; Score 13; DB 9; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.8e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 gctgctggtggaa 20
|||||
Db 20 GCTGCTGCTGGAA 8

RESULT 10
AX044138/c 23 bp DNA PAT 24-NOV-2000
LOCUS AX044138
DEFINITION Sequence 38 from Patent WO0066747.
ACCESSION AX044138
VERSION AX044138.1 GI:11343016
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 23)
AUTHORS Hawkes, T.R., Warner, S.A., Andrews, C.J., Bachoo, S. and
Pickerill, A.P.
TITLE Herbicide resistant plants
JOURNAL Patent: WO 0066747-A 38 09-NOV-2000;
ZENECA LIMITED (GB)
FEATURES
source Location/Qualifiers
1..23
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer"
BASE COUNT 5 a 7 c 5 g 6 t
ORIGIN
Query Match 65.0%; Score 13; DB 9; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.8e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 gctgctggtggaa 20
|||||
Db 20 GCTGCTGCTGGAA 8

RESULT 11
AX044178/c 23 bp DNA PAT 24-NOV-2000
LOCUS AX044178
DEFINITION Sequence 30 from Patent WO0066746.
ACCESSION AX044178
VERSION AX044178.1 GI:11343056
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 23)
AUTHORS Hawkes, T.R., Warner, S.A., Andrews, C.J., Bachoo, S. and
Pickerill, A.P.
TITLE Herbicide resistant plants
JOURNAL Patent: WO 0066746-A 30 09-NOV-2000;
ZENECA LIMITED (GB)
FEATURES
source Location/Qualifiers
1..23
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer"
BASE COUNT 5 a 7 c 5 g 6 t
ORIGIN

```

Query Match 65.0%; Score 13; DB 9; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 gctgctggtgaa 20  
 |||||  
 Db 20 GCTGCTGCTGGA 8

RESULT 12  
 HSA224214/c 48 bp mRNA PRI 23-FEB-2001  
 LOCUS Homo sapiens mRNA for T cell receptor beta chain V-D-J junctional  
 DEFINITION region (TCRBV6B1J55).  
 ACCESSION AJ224214  
 VERSION AJ224214.1 GI:13159943  
 KEYWORDS T cell receptor; T cell receptor beta chain; variable region.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 48)  
 AUTHORS Silins,S.L., Cross,S.M., Krauer,K.G., Moss,D.J., Schmidt,C.W. and Misko,I.S.  
 TITLE A functional link for major TCR expansions in healthy adults caused by persistent Epstein-Barr virus infection  
 JOURNAL J Clin. Invest. 102 (8), 1551-1558 (1998)  
 MEDLINE 99007392  
 REFERENCE 2 (bases 1 to 48)  
 AUTHORS Silins,S.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-FEB-1998) Silins S.L., Queensland Institute of Medical Research, The Bancroft Centre, 300 Herston Road, Brisbane, AUSTRALIA 4029

FEATURES  
 source Location/Qualifiers  
 1. .48  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /haplotype="A1, A3, B8, B35"  
 /cell\_line="JS11"  
 /cell\_type="cytotoxic T lymphocyte"  
 /tissue\_type="blood"  
 /rearranged

V.region 1. .48  
 BASE COUNT 7 a 13 c 18 g 10 t  
 ORIGIN

Query Match 65.0%; Score 13; DB 92; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tctcgcgtgtgg 15  
 |||||  
 Db 17 TCCTCGCTGCTGG 5

RESULT 13  
 AR103323/c 20 bp DNA PAT 14-FEB-2001  
 LOCUS Sequence 43 from patent US 6087173.  
 ACCESSION AR103323  
 VERSION AR103323.1 GI:12814911  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Bennett,C.Frank, Ackermann,E.J. and Cowser,L.M.  
 TITLE Antisense modulation of X-linked inhibitor of apoptosis expression  
 JOURNAL Patent: US 6087173-A 43 11-JUL-2000;

FEATURES  
 source Location/Qualifiers  
 1. .20  
 /organism="unknown"  
 BASE COUNT 7 a 6 c 3 g 4 t  
 ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 20;  
 Best Local Similarity 78.9%; Pred. No. 7.5e+05;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attctcgcgtgtgga 19  
 |||||  
 Db 19 ATTGTTCCATGCTGCTGGA 1

RESULT 14  
 I25076 30 bp DNA PAT 07-OCT-1996  
 LOCUS Sequence 3 from patent US 5547864.  
 DEFINITION I25076  
 ACCESSION I25076  
 VERSION I25076.1 GI:1604946  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 30)  
 AUTHORS Kawasaki,H., Tsuchiya,M., Miwa,K. and Kawahara,Y.  
 TITLE Coryneform bacteria deficient in a cell surface protein  
 JOURNAL Patent: US 5547864-A 3 20-AUG-1996;  
 FEATURES Location/Qualifiers  
 1. .30  
 source /organism="unknown"  
 BASE COUNT 7 a 9 c 7 g 7 t  
 ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 30;  
 Best Local Similarity 78.9%; Pred. No. 6.6e+05;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttctcgcgtgtgga 20  
 |||||  
 Db 1 TTCATCGCTGCTGCGCA 19

RESULT 15  
 I71129 30 bp DNA PAT 03-APR-1998  
 LOCUS Sequence 3 from patent US 5681717.  
 ACCESSION I71129  
 VERSION I71129.1 GI:3007264  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 30)  
 AUTHORS Kawasaki,H., Tsuchiya,M., Miwa,K. and Kawahara,Y.  
 TITLE DNA encoding novel cell surface protein  
 JOURNAL Patent: US 5681717-A 3 28-OCT-1997;  
 FEATURES Location/Qualifiers  
 1. .30  
 source /organism="unknown"  
 BASE COUNT 7 a 9 c 7 g 7 t  
 ORIGIN

Query Match 63.0%; Score 12.6; DB 10; Length 30;  
 Best Local Similarity 78.9%; Pred. No. 6.6e+05;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttctcgcgtgtgga 20  
 |||||

Db 1 TTCATCGCTGCTGTCGGCA 19

Search completed: October 2, 2001, 15:56:50  
Job time: 14174 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:06 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-17

Perfect score: 20

Sequence: 1 attctcgtcgtggtggaa 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est33.\*
- 26: gb\_est34.\*
- 27: gb\_est35.\*
- 28: gb\_est36.\*
- 29: gb\_est37.\*
- 30: gb\_est38.\*
- 31: gb\_est39.\*
- 32: gb\_est40.\*
- 33: em\_estba.\*
- 34: em\_estfun.\*
- 35: em\_esthum1.\*
- 36: em\_esthum2.\*
- 37: em\_esthum3.\*
- 38: em\_esthum4.\*
- 39: em\_esthum5.\*
- 40: em\_esthum6.\*
- 41: em\_esthum7.\*
- 42: em\_esthum8.\*
- 43: em\_esthum9.\*

- 44: em\_esthum10.\*
- 45: em\_esthum11.\*
- 46: em\_esthum12.\*
- 47: em\_esthum13.\*
- 48: em\_esthum14.\*
- 49: em\_esthum15.\*
- 50: em\_esthum16.\*
- 51: em\_esthum17.\*
- 52: em\_esthum18.\*
- 53: em\_esthum19.\*
- 54: em\_esthum20.\*
- 55: em\_esthum21.\*
- 56: em\_esthum22.\*
- 57: em\_esthum23.\*
- 58: em\_esthum24.\*
- 59: em\_esthum25.\*
- 60: em\_esthum26.\*
- 61: em\_esthum27.\*
- 62: em\_esthum28.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estin5.\*
- 68: em\_estom1.\*
- 69: em\_estom2.\*
- 70: em\_estov1.\*
- 71: em\_estov2.\*
- 72: em\_estpl1.\*
- 73: em\_estpl2.\*
- 74: em\_estpl3.\*
- 75: em\_estpl4.\*
- 76: em\_estpl5.\*
- 77: em\_estpl6.\*
- 78: em\_estpl7.\*
- 79: em\_estpl8.\*
- 80: em\_estpl9.\*
- 81: em\_estpl10.\*
- 82: em\_estrol.\*
- 83: em\_estrol2.\*
- 84: em\_estrol3.\*
- 85: em\_estrol4.\*
- 86: em\_estrol5.\*
- 87: em\_estrol6.\*
- 88: em\_estrol7.\*
- 89: em\_estrol8.\*
- 90: em\_estrol9.\*
- 91: em\_estrol10.\*
- 92: em\_estrol11.\*
- 93: em\_estrol12.\*
- 94: em\_estrol13.\*
- 95: em\_estrol14.\*
- 96: em\_estrol15.\*
- 97: em\_estrol16.\*
- 98: em\_estrol17.\*
- 99: em\_estrol18.\*
- 100: em\_estrol19.\*
- 101: em\_estrol20.\*
- 102: gb\_est25.\*
- 103: gb\_est26.\*
- 104: gb\_est27.\*
- 105: gb\_est28.\*
- 106: gb\_est29.\*
- 107: gb\_est30.\*
- 108: gb\_est31.\*
- 109: gb\_est32.\*
- 110: gb\_est41.\*
- 111: gb\_est42.\*
- 112: gb\_est43.\*
- 113: gb\_est44.\*
- 114: gb\_est45.\*
- 115: gb\_est46.\*
- 116: gb\_est47.\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	15.2	76.0	37	11	AA734467 vt94a08.r
C 2	13.2	66.0	32	244	AZ473098 1M0288114
C 3	13	65.0	50	107	AU107725 AU107725
C 4	13	65.0	50	107	AU107726 AU107726
C 5	13	65.0	50	107	AU107729 AU107729
C 6	13	65.0	50	107	AU107730 AU107730
C 7	13	65.0	50	107	AU107732 AU107732
C 8	13	65.0	50	107	AU107733 AU107733
C 9	13	65.0	50	107	AU107734 AU107734
C 10	13	65.0	50	107	AU107736 AU107736
C 11	13	65.0	50	107	AU107737 AU107737
C 12	13	65.0	50	107	AU107738 AU107738
C 13	13	65.0	50	107	AU107739 AU107739
C 14	12.8	64.0	45	246	AZ599531 1M0414M23
C 15	12.8	64.0	50	107	AU105903 AU105903
C 16	12.6	63.0	42	249	AZ771500 1M0573M12
C 17	12.6	63.0	50	107	AU102691 AU102691
C 18	12.4	62.0	40	250	AZ835855 2M0130K18
C 19	12.4	62.0	49	11	AA780186 af46f07.s
C 20	12.2	61.0	30	250	AZ821583 2M0094I23
C 21	12.2	61.0	50	107	AU106715 AU106715
C 22	12.2	61.0	50	107	AU106716 AU106716
C 23	12.2	61.0	50	107	AU106717 AU106717
C 24	12.2	61.0	50	107	AU106718 AU106718
C 25	12.2	61.0	50	107	AU106719 AU106719
C 26	12.2	61.0	50	107	AU106722 AU106722
C 27	12.2	61.0	50	107	AU106723 AU106723
C 28	12.2	61.0	50	107	AU106724 AU106724
C 29	12.2	61.0	50	107	AU106725 AU106725
C 30	12.2	61.0	50	107	AU106726 AU106726
C 31	12.2	61.0	50	107	AU106729 AU106729
C 32	12.2	61.0	50	107	AU106733 AU106733
C 33	12.2	61.0	50	107	AU106734 AU106734
C 34	12.2	61.0	50	107	AU106735 AU106735
C 35	12	60.0	50	107	AU104880 AU104880
C 36	12	60.0	50	249	AZ802269 2M0061C05
C 37	11.8	59.0	34	23	AI657872 fc14f02.y
C 38	11.8	59.0	41	223	AQ025523 Ep(X)1594
C 39	11.8	59.0	44	187	R35634 y966h08.r1
C 40	11.8	59.0	50	107	AU103987 AU103987
C 41	11.8	59.0	50	107	AU103994 AU103994
C 42	11.6	58.0	23	244	AZ465057 1M0274D15
C 43	11.6	58.0	34	141	BE914450 601668B34
C 44	11.6	58.0	40	9	AA618222 nq16a07.s
C 45	11.6	58.0	46	244	AZ484138 1M0310E20

## ALIGNMENTS

RESULT 1	AA734467/c	37 bp	EST	07-JAN-1998
LOCUS	vt94a08.r1 Soares_mammary_gland_NbMMG			
DEFINITION	IMAGE:1178774 5' similar to TR:Q09472 Q09472 E1A-ASSOCIATED PROTEIN P300. ; mRNA sequence.			
ACCESSION	AA734467			
VERSION	AA734467.1	GI:2756134		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 37)			
	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, B., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,			

<p>Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.</p> <p>The WashU-HMMI Mouse EST Project</p> <p>Unpublished (1996)</p> <p>Contact: Marra M/Mouse EST Project</p> <p>WashU-HMMI Mouse EST Project</p> <p>Washington University School of Medicine</p> <p>4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108</p> <p>Tel: 314 286 1800</p> <p>Fax: 314 286 1810</p> <p>Email: mouseest@wustl.edu</p> <p>This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.</p> <p>MG1:636622</p> <p>Trace considered overall poor quality</p> <p>Possible reversed clone: similarity on wrong strand</p> <p>Seq primer: -28m13 rev2 ET from Amersham</p> <p>High quality sequence stop: 1.</p> <p>Location/Qualifiers</p> <p>1. .37</p> <p>/organism="Mus musculus"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="IMAGE:1178774"</p> <p>/clone_lib="Soares_mammary_gland_NbMMG"</p> <p>/sex="male"</p> <p>/tissue_type="mammary gland"</p> <p>/dev_stage="4 weeks"</p> <p>/lab_host="DH10B"</p> <p>/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAATGGAGGGCGGCAATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."</p> <p>BASE COUNT 10 a 14 c 7 g 6 t</p> <p>ORIGIN</p> <p>Query Match 76.0% Score 15.2; DB 11; Length 37;</p> <p>Best Local Similarity 85.0%; Pred. No. 4.6e+03;</p> <p>Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p> <p>Qy 1 attctcgtcgtcgttgga 20</p> <p>                 </p> <p>Db 24 ATTCTTGGTCTGCTGGAA 5</p>	<p>RESULT 2</p> <p>AZ473098</p> <p>LOCUS</p> <p>DEFINITION</p> <p>1M0288114R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0288114 R, DNA sequence.</p> <p>ACCESSION</p> <p>AZ473098</p> <p>VERSION</p> <p>AZ473098.1</p> <p>KEYWORDS</p> <p>GSS.</p> <p>SOURCE</p> <p>house mouse.</p> <p>ORGANISM</p> <p>Mus musculus</p> <p>Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;</p> <p>Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 32)</p> <p>Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.</p> <p>Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts</p> <p>Unpublished (2000)</p>
--	---



```

RESULT 5
AU107729/c
LOCUS AU107729 50 bp mRNA EST 05-APR-2001
DEFINITION Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HEP01448, mRNA sequence.
ACCESSION AU107729
VERSION AU107729.1 GI:13557250
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 50)
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
,K., Suyama,A. and Sugano,S.
TITLE Fine Structural analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL Unpublished (2001)
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., and Sugano
,S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES
source Location/Qualifiers
1..50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP01448"
/clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT 8 a 21 c 13 g 8 t
ORIGIN
Query Match 65.0%; Score 13; DB 107; Length 50;
Best Local Similarity 100.0%; Pred. No. 4.9e+04;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 cgctgctggtgga 19
|||||
Db 49 CGCTGCTGGTGGG 37

RESULT 6
AU107730/c
LOCUS AU107730 50 bp mRNA EST 05-APR-2001
DEFINITION Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HEP01701, mRNA sequence.
ACCESSION AU107730
VERSION AU107730.1 GI:13557251
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 50)
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
,K., Suyama,A. and Sugano,S.
TITLE Fine Structural analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL Unpublished (2001)
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., and Sugano
,S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

```

```

FEATURES
source Location/Qualifiers
1..50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP01701"
/clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT 8 a 23 c 13 g 6 t
ORIGIN
Query Match 65.0%; Score 13; DB 107; Length 50;
Best Local Similarity 100.0%; Pred. No. 4.9e+04;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 cgctgctggtgga 19
|||||
Db 44 CGCTGCTGGTGGG 32

RESULT 7
AU107732/c
LOCUS AU107732 50 bp mRNA EST 05-APR-2001
DEFINITION Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HEP13686, mRNA sequence.
ACCESSION AU107732
VERSION AU107732.1 GI:13557253
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 50)
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
,K., Suyama,A. and Sugano,S.
TITLE Fine Structural analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
JOURNAL Unpublished (2001)
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., and Sugano
,S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES
source Location/Qualifiers
1..50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP13686"
/clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT 8 a 22 c 14 g 6 t
ORIGIN
Query Match 65.0%; Score 13; DB 107; Length 50;
Best Local Similarity 100.0%; Pred. No. 4.9e+04;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 cgctgctggtgga 19
|||||
Db 43 CGCTGCTGGTGGG 31

RESULT 8
AU107733/c
LOCUS AU107733 50 bp mRNA EST 05-APR-2001
DEFINITION Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HEP13813, mRNA sequence.
ACCESSION AU107733
VERSION AU107733.1 GI:13557254
KEYWORDS EST.

```

SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
,K., Suyama,A. and Sugano,S.  
TITLE Fine Structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
JOURNAL Unpublished (2001)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
,S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  
FEATURES  
source 1..50  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP13813"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 8 a 23 c 13 g 6 t  
ORIGIN

Query Match 65.0%; Score 13; DB 107; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.9e+04;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 7 cgcctcgtggtgga 19  
|||||  
Db 45 CGCTCTGGTGA 33

RESULT 9  
AUI07734/c  
LOCUS AUI07734 50 bp mRNA EST 05-APR-2001  
DEFINITION AUI07734 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
HEP14963, mRNA sequence.  
ACCESSION AUI07734  
VERSION AUI07734.1 GI:13557255  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
,K., Suyama,A. and Sugano,S.  
TITLE Fine Structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
JOURNAL Unpublished (2001)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
,S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  
FEATURES  
source 1..50  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP14963"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 8 a 20 c 13 g 9 t  
ORIGIN

Query Match 65.0%; Score 13; DB 107; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.9e+04;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 7 cgcctcgtggtgga 19  
|||||  
Db 43 CGCTCTGGTGA 31

RESULT 10  
AUI07736/c  
LOCUS AUI07736 50 bp mRNA EST 05-APR-2001  
DEFINITION AUI07736 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
KAT09254, mRNA sequence.  
ACCESSION AUI07736  
VERSION AUI07736.1 GI:13557257  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
,K., Suyama,A. and Sugano,S.  
TITLE Fine Structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
JOURNAL Unpublished (2001)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
,S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  
FEATURES  
source 1..50  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="KAT09254"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 7 a 20 c 14 g 9 t  
ORIGIN

Query Match 65.0%; Score 13; DB 107; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.9e+04;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 7 cgcctcgtggtgga 19  
|||||  
Db 39 CGCTCTGGTGA 27

RESULT 11  
AUI07737/c  
LOCUS AUI07737 50 bp mRNA EST 05-APR-2001  
DEFINITION AUI07737 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
REC00765, mRNA sequence.  
ACCESSION AUI07737  
VERSION AUI07737.1 GI:13557258  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
,K., Suyama,A. and Sugano,S.

**TITLE** Fine Structural analysis of transcription start sites of human  
**JOURNAL** mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
**COMMENT** Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, A. and Sugano  
 S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

**FEATURES**

source  
 1. .50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="REC00765"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
 8 a 22 c 14 g 6 t

**BASE COUNT**  
**ORIGIN**

Query Match 65.0%; Score 13; DB 107; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 4.9e+04;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 cgctgctggtgga 19  
 |||||

Db 43 CGCTGCTGGTGA 31

**RESULT 12**  
**AU107738/c**

**LOCUS** AU107738 50 bp mRNA EST 05-APR-2001  
**DEFINITION** AU107738 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 2RV61429, mRNA sequence.

**ACCESSION** AU107738  
**VERSION** AU107738.1 GI:13557259  
**KEYWORDS** EST.  
**SOURCE** human.

**ORGANISM**

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**  
**AUTHORS** Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata  
 H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo  
 K., Suyama, A. and Sugano, S.

**TITLE** Fine Structural analysis of transcription start sites of human  
**JOURNAL** mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
**COMMENT** Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, A. and Sugano  
 S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

**FEATURES**

source  
 1. .50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2RV61429"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
 7 a 22 c 15 g 6 t

**BASE COUNT**  
**ORIGIN**

Query Match 65.0%; Score 13; DB 107; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 4.9e+04;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 cgctgctggtgga 19  
 |||||

Db 39 CGCTGCTGGTGA 27

**RESULT 13**  
**AU107739/c**

**LOCUS** AU107739 50 bp mRNA EST 05-APR-2001  
**DEFINITION** AU107739 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 2RV61978, mRNA sequence.

**ACCESSION** AU107739  
**VERSION** AU107739.1 GI:13557260  
**KEYWORDS** EST.  
**SOURCE** human.

**ORGANISM**

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**  
**AUTHORS** Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata  
 H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo  
 K., Suyama, A. and Sugano, S.

**TITLE** Fine Structural analysis of transcription start sites of human  
**JOURNAL** mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
**COMMENT** Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, A. and Sugano  
 S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

**FEATURES**  
 source  
 1. .50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2RV61978"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
 7 a 22 c 15 g 6 t

**BASE COUNT**  
**ORIGIN**

Query Match 65.0%; Score 13; DB 107; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 4.9e+04;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 cgctgctggtgga 19  
 |||||

Db 39 CGCTGCTGGTGA 27

**RESULT 14**  
**AZ599531**

**LOCUS** AZ599531 45 bp DNA GSS 13-DEC-2000  
**DEFINITION** IM0414M23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0414M23 R, DNA sequence.

**ACCESSION** AZ599531  
**VERSION** AZ599531.1 GI:11721721  
**KEYWORDS** GSS.  
**SOURCE** house mouse.

**ORGANISM**

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 45)

**REFERENCE**  
**AUTHORS** Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.  
**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0414 row: M column: 23

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 45.

## FEATURES

Location/Qualifiers

source

1. .45

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGCIM0414M23"

/clone\_lib="Mouse 10kb plasmid UUGCIM library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (gil4732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

6 a 10 c 8 g 21 t

BASE COUNT  
ORIGIN

Query Match 64.0%; Score 12.8; DB 246; Length 45;

Best Local Similarity 87.5%; Pred. No. 6.1e+04;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ttctctgcgtcgtggtg 17

||||| ||||| |||||

Db 20 TTCCTGCTGCTGTGTG 35

## RESULT 15

AU105903

LOCUS

AU105903 50 bp mRNA EST 05-APR-2001

DEFINITION AU105903 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone

KAT08022, mRNA sequence.

ACCESSION AU105903

VERSION AU105903.1 GI:13555424

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 50)

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata

,K., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo

H., Suyama,A. and Sugano,S.

Fine Structural analysis of transcription start sites of human

mRNAs using full-length enriched and 5'-end enriched cDNA libraries

Unpublished (2001)

JOURNAL Contact: Yutaka Suzuki

COMMENT Department of Virology

Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: ysuzuki@ims.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

Location/Qualifiers

source

1. .50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="KAT08022"

/clone\_lib="Sugano Homo sapiens cDNA library"

3 a 21 c 9 g 17 t

Query Match 64.0%;

Best Local Similarity 87.5%; Score 12.8; DB 107; Length 50;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ttctctgcgtcgtggtg 17

||||| ||||| |||||

Db 13 TTCGTCGCTGCTGCTG 28

Search completed: October 2, 2001, 15:01:07

Job time: 10836 sec



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:51 ; Search time 3339.34 Seconds  
(without alignments)  
92.640 Million cell updates/sec

Title: US-09-757-100B-20  
Perfect score: 20  
Sequence: 1 catgatgcttaaaagcttac 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htgo\_hum1:\*  
23: em\_htgo\_hum2:\*  
24: em\_htgo\_hum3:\*  
25: em\_htgo\_hum4:\*  
26: em\_htgo\_hum5:\*  
27: em\_htgo\_hum6:\*  
28: em\_htgo\_hum7:\*  
29: em\_htgo\_hum8:\*  
30: em\_htgo\_inv1:\*  
31: em\_htgo\_inv2:\*  
32: em\_htgo\_other:\*  
33: em\_hum1\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vil:\*  
59: gb\_vil2:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_rol:\*  
95: gb\_rol2:\*  
96: gb\_in4:\*  
97: gb\_pri0:\*  
98: em\_ba3:\*

pred, No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	13.8	69.0	29	9	AR038139	AR038139 Sequence
C 2	13.8	69.0	29	10	I79867	I79867 Sequence 7
C 3	13.6	68.0	48	9	A87694	A87694 Sequence 2
C 4	13.4	67.0	44	9	AR075944	AR075944 Sequence
C 5	13.2	66.0	20	45	E10769	E10769 PCR primer
C 6	13.2	66.0	20	45	E10798	E10798 PCR primer
C 7	13.2	66.0	33	9	A97414	A97414 Sequence 4
C 8	13.2	66.0	33	9	AR022050	AR022050 Sequence

c 9 13.2 66.0 33 10 I92840 Sequence 14  
 c 10 12.6 63.0 35 9 A61915 Sequence 3  
 c 11 12.2 61.0 25 9 AR049871 Sequence  
 c 12 12.2 61.0 25 10 I59435 Sequence 10  
 c 13 12.2 61.0 27 9 AR110319 Sequence  
 c 14 12.2 61.0 29 9 A65301 Sequence 24  
 c 15 12.2 61.0 29 9 AR036134 Sequence  
 c 16 12.2 61.0 29 9 AR069509 Sequence  
 c 17 12.2 61.0 29 9 AR076177 Sequence  
 c 18 12.2 61.0 30 10 I77055 Sequence 2  
 c 19 12.2 61.0 40 9 A65366 Sequence 2  
 c 20 12 60.0 30 10 AX100542 Sequence  
 c 21 12 60.0 44 9 AR071481 Sequence  
 c 22 12 60.0 44 9 AR080501 Sequence  
 c 23 11.8 59.0 21 9 AR071089 Sequence  
 c 24 11.8 59.0 30 9 AR063799 Sequence  
 c 25 11.8 59.0 30 10 I30162 Sequence 84  
 c 26 11.8 59.0 34 9 AR079838 Sequence  
 c 27 11.8 59.0 34 53 C75779 Sequence  
 c 28 11.8 59.0 37 9 AR028960 Sequence  
 c 29 11.8 59.0 45 9 A23344 Artificial  
 c 30 11.8 59.0 45 9 AR068522 Sequence  
 c 31 11.8 59.0 47 54 G54572 Sequence  
 c 32 11.6 58.0 21 10 E27054 Novel recep  
 c 33 11.6 58.0 23 9 AR079557 Sequence  
 c 34 11.6 58.0 25 9 AR075352 Sequence  
 c 35 11.6 58.0 26 9 A48469 Sequence 8  
 c 36 11.6 58.0 26 9 AR050991 Sequence  
 c 37 11.6 58.0 26 9 AR088908 Sequence  
 c 38 11.6 58.0 26 9 AX000672 Sequence  
 c 39 11.6 58.0 30 10 I23896 Sequence 13  
 c 40 11.6 58.0 32 10 I17561 Sequence 24  
 c 41 11.6 58.0 33 10 AX084116 Sequence  
 c 42 11.6 58.0 33 10 E15443 PCR primer  
 c 43 11.6 58.0 35 9 A21585 Sequence  
 c 44 11.6 58.0 38 9 AR084169 Sequence  
 c 45 11.6 58.0 40 9 A36486 Sequence 27

## ALIGNMENTS

RESULT 1  
 AR038139/c  
 LOCUS AR038139 29 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 9 from patent US 5804417.  
 ACCESSION AR038139  
 VERSION AR038139.1 GI:5956856  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 29)  
 AUTHORS Martens,G.JulianusMaria, Chaudhuri,B. and Stephan,C.  
 TITLE Recombinant production of proteins using 7B2 protein  
 JOURNAL Patent: US 5804417-A 9 08-SEP-1998;  
 FEATURES Location/Qualifiers  
 source 1..29  
 BASE COUNT 11 a 6 c 3 g 9 t  
 ORIGIN  
 Query Match 69.0%; Score 13.8; DB 9; Length 29;  
 Best Local Similarity 88.2%; Pred. No. 1.6e+04;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 3 tgatgcttaaaagctta 19  
 ||||| ||||| |||||  
 Db 18 TGTTCATAAAAGCTTA 2  
 RESULT 2  
 AR075944  
 LOCUS AR075944 44 bp DNA PAT 30-AUG-2000  
 DEFINITION Sequence 5 from patent US 5958713.  
 ACCESSION AR075944  
 VERSION AR075944.1 GI:10002690  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 48)  
 AUTHORS Drevet,P. and Menez,A.  
 TITLE NON-COVALENT COMPLEX COMPRISING AT LEAST AN ANTIBODY AND ELEMENT  
 BINDING WITH IMMUNOGLOBULINS ASSOCIATED WITH AN ACTIVE SUBSTANCE,  
 METHOD OF PREPARING AND APPLICATIONS  
 JOURNAL Patent: WO 9834956-A 2 13-AUG-1998;  
 FEATURES Location/Qualifiers  
 source 1..48  
 BASE COUNT 14 a 9 c 10 g 15 t  
 ORIGIN  
 Query Match 68.0%; Score 13.6; DB 9; Length 48;  
 Best Local Similarity 80.0%; Pred. No. 2.1e+04;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 catgactcttaaaagcttac 20  
 ||||| ||||| ||||| |||||  
 Db 8 CAGGATCCTTATAAGCTTCC 27  
 RESULT 4  
 AR075944  
 LOCUS AR075944 44 bp DNA PAT 30-AUG-2000  
 DEFINITION Sequence 5 from patent US 5958713.  
 ACCESSION AR075944  
 VERSION AR075944.1 GI:10002690  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 48)  
 AUTHORS Drevet,P. and Menez,A.  
 TITLE NON-COVALENT COMPLEX COMPRISING AT LEAST AN ANTIBODY AND ELEMENT  
 BINDING WITH IMMUNOGLOBULINS ASSOCIATED WITH AN ACTIVE SUBSTANCE,  
 METHOD OF PREPARING AND APPLICATIONS  
 JOURNAL Patent: WO 9834956-A 2 13-AUG-1998;  
 FEATURES Location/Qualifiers  
 source 1..48  
 BASE COUNT 14 a 9 c 10 g 15 t  
 ORIGIN

I79867/c  
 LOCUS I79867 29 bp DNA PAT 10-JUN-1998  
 DEFINITION Sequence 7 from patent US 5708140.  
 ACCESSION I79867  
 VERSION I79867.1 GI:3208157  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 29)  
 AUTHORS Martens,G.JulianusMaria, Chaudhuri,B. and Stephan,C.  
 TITLE Production of proteins using 7B2 protein  
 JOURNAL Patent: US 5708140-A 7 13-JAN-1998;  
 FEATURES Location/Qualifiers  
 source 1..29  
 BASE COUNT 11 a 6 c 3 g 9 t  
 ORIGIN

Query Match 69.0%; Score 13.8; DB 10; Length 29;  
 Best Local Similarity 88.2%; Pred. No. 1.6e+04;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 tgatgcttaaaagctta 19  
 ||||| ||||| |||||  
 Db 18 TGTTCATAAAAGCTTA 2

RESULT 3  
 AR7694  
 LOCUS AR7694 48 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 2 from Patent WO9834956.  
 ACCESSION AR7694  
 VERSION AR7694.1 GI:6736312  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 48)  
 AUTHORS Drevet,P. and Menez,A.  
 TITLE NON-COVALENT COMPLEX COMPRISING AT LEAST AN ANTIBODY AND ELEMENT  
 BINDING WITH IMMUNOGLOBULINS ASSOCIATED WITH AN ACTIVE SUBSTANCE,  
 METHOD OF PREPARING AND APPLICATIONS  
 JOURNAL Patent: WO 9834956-A 2 13-AUG-1998;  
 FEATURES Location/Qualifiers  
 source 1..48  
 BASE COUNT 14 a 9 c 10 g 15 t  
 ORIGIN

Query Match 68.0%; Score 13.6; DB 9; Length 48;  
 Best Local Similarity 80.0%; Pred. No. 2.1e+04;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 catgactcttaaaagcttac 20  
 ||||| ||||| ||||| |||||  
 Db 8 CAGGATCCTTATAAGCTTCC 27

RESULT 4  
 AR075944  
 LOCUS AR075944 44 bp DNA PAT 30-AUG-2000  
 DEFINITION Sequence 5 from patent US 5958713.  
 ACCESSION AR075944  
 VERSION AR075944.1 GI:10002690  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 48)  
 AUTHORS Drevet,P. and Menez,A.  
 TITLE NON-COVALENT COMPLEX COMPRISING AT LEAST AN ANTIBODY AND ELEMENT  
 BINDING WITH IMMUNOGLOBULINS ASSOCIATED WITH AN ACTIVE SUBSTANCE,  
 METHOD OF PREPARING AND APPLICATIONS  
 JOURNAL Patent: WO 9834956-A 2 13-AUG-1998;  
 FEATURES Location/Qualifiers  
 source 1..48  
 BASE COUNT 14 a 9 c 10 g 15 t  
 ORIGIN

Query Match 68.0%; Score 13.6; DB 9; Length 48;  
 Best Local Similarity 80.0%; Pred. No. 2.1e+04;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 catgactcttaaaagcttac 20  
 ||||| ||||| ||||| |||||  
 Db 8 CAGGATCCTTATAAGCTTCC 27

REFERENCE 1 (bases 1 to 44)  
AUTHORS Thastrup,O., Tullin,S. slashed.ren, Poulsen,L.Kongsbak and Bj.o slashed.rm,S.petersen.  
TITLE Method of detecting biologically active substances by using green fluorescent protein  
JOURNAL Patent: US 5958713-A 5 28-SEP-1999;  
FEATURES Location/Qualifiers  
source  
1. .44

BASE COUNT 12 a 8 c 11 g 13 t  
ORIGIN

Query Match 67.0%; Score 13.4; DB 9; Length 44;  
Best Local Similarity 93.3%; Pred. No. 2.6e+04;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 atgcttaaagctta 19  
II IIIIIIIIIII

Db 5 ATTCTTAAAGCTTA 19

RESULT 5  
E10769/c standard; DNA; UNC; 20 BP.  
ID E10769  
XX  
AC E10769;  
XX  
SV E10769.1

XX  
DT 08-OCT-1997 (Rel. 52, Created)  
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)  
XX  
DE PCR primer for gaining human serum albumin.  
XX  
KW JP 1996051982-A/11.

XX  
OS unidentifed  
OC unclassified.  
XX  
RN [1]

RP 1-20  
RA Higashida H., Murakami K., Hama Y., Tsukamoto Y., Isoai A., Kumagai H.;  
RT "MODIFIED GENE CODING HUMAN SERUM ALBUMEN";  
RL Patent number JP1996051982-A/11, 27-FEB-1996.  
RL ASAHI GLASS CO LTD.

XX  
CC OS None  
CC OC Artificial sequences.  
CC PN JP 1996051982-A/11  
CC PD 27-FEB-1996  
CC PF 11-AUG-1994 JP 1994209369  
CC PI HIGASHIDA HIDEKI, MURAKAMI KIMIKO, HAMA YUKO, TSUKAMOTO YOKO,  
CC FI ISOAI ATSUSHI, KUMAGAI HIROMICHI  
CC PC C12N15/09,C07K19/00//C12N1/19,C12P21/02,(C12N1/19,C12R1:645),  
CC PC (C12P21/02,  
CC PC C12R1:645);  
CC CC strandedness: Single;  
CC CC topology: Linear;  
CC CC hypothetical: No;  
CC CC anti-sense: Yes;

CC FH Key Location/Qualifiers

CC FT source 1. .20 /organism="Artificial sequences"

XX Key Location/Qualifiers

FH source 1. .20 /db\_xref="taxon:32644"  
FT /organism="unidentified"

XX  
SQ Sequence 20 BP; 6 A; 3 C; 2 G; 9 T; 0 other;

Query Match 66.0%; Score 13.2; DB 45; Length 20;  
Best Local Similarity 83.3%; Pred. No. 3.2e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 atgatgcttaaagctta 19  
IIII I IIIIIIIII

Db 20 ATGAAGATCAAAAGCTTA 3

RESULT 6  
E10798/c standard; DNA; UNC; 20 BP.  
ID E10798  
XX  
AC E10798;

XX  
SV E10798.1

XX  
DT 08-OCT-1997 (Rel. 52, Created)  
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)  
XX  
DE PCR primer for gaining human serum albumin.  
XX  
KW JP 1996053500-A/8.

XX  
OS unidentifed  
OC unclassified.  
XX  
RN [1]

RP 1-20  
RA Higashida H., Murakami K., Hama Y., Tsukamoto Y., Isoai A., Kumagai H.;  
RT "FUSED PROTEIN AND GENE CODING SAID PROTEIN";  
RL Patent number JP1996053500-A/8, 27-FEB-1996.  
RL ASAHI GLASS CO LTD.

XX  
OS None  
CC OC Artificial sequences.  
CC PN JP 1996053500-A/8  
CC PD 27-FEB-1996  
CC PF 11-AUG-1994 JP 1994209368

CC PI HIGASHIDA HIDEKI, MURAKAMI KIMIKO, HAMA YUKO, TSUKAMOTO YOKO,  
CC FI ISOAI ATSUSHI, KUMAGAI HIROMICHI  
CC PC C07K19/00,C12N1/19,C12N15/09,C12P21/02,(C12N1/19,C12R1:645),  
CC PC (C12P21/02,  
CC PC C12R1:645);  
CC CC strandedness: Single;  
CC CC topology: Linear;

CC FH Key Location/Qualifiers

CC FT source 1. .20 /organism="Artificial sequences"

XX Key Location/Qualifiers

FH source 1. .20 /db\_xref="taxon:32644"  
FT /organism="unidentified"

XX  
SQ Sequence 20 BP; 6 A; 3 C; 2 G; 9 T; 0 other;

Query Match 66.0%; Score 13.2; DB 45; Length 20;  
Best Local Similarity 83.3%; Pred. No. 3.2e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 atgatgcttaaagctta 19  
IIII I IIIIIIIII

Db 20 ATGAAGATCAAAAGCTTA 3

RESULT 7  
A97414

LOCUS A97414 33 bp DNA 26-JAN-2000  
 DEFINITION Sequence 4 from Patent WO9916867.  
 ACCESSION A97414 PAT  
 VERSION A97414.1 GI:6780760  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 33)  
 AUTHORS Waldmann H. and Zelenika, D.  
 TITLE TRANSGENIC MODEL COMPRISING TCR ALPHA AND BETA CHAINS  
 JOURNAL Patent: WO 9916867-A 4 08-APR-1999;  
 COBBOLD STEPHEN P (GB); ISIS INNOVATION (GB)  
 FEATURES  
 source  
 1. .33  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 10 a 5 c 9 g 9 t  
 ORIGIN

Query Match 66.0%; Score 13.2; DB 9; Length 33;  
 Best Local Similarity 83.3%; Pred. No. 3.3e+04;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 catgatgcttaaaagctt 18  
 || ||||| ||||| ||  
 Db 8 CAGGATGCATAAAGTTT 25

RESULT 8  
 AR022050/c  
 LOCUS AR022050 33 bp DNA 05-DEC-1998  
 DEFINITION Sequence 14 from patent US 5792456.  
 ACCESSION AR022050 PAT  
 VERSION AR022050.1 GI:3976112  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 33)  
 AUTHORS Yelton, D., Glaser, S., Huse, W. and Rosok, M. Joanne.  
 TITLE Mutant BR96 antibodies reactive with human carcinomas  
 JOURNAL Patent: US 5792456-A 14 11-AUG-1998;  
 FEATURES  
 source  
 1. .33  
 /organism="unknown"  
 BASE COUNT 12 a 7 c 7 g 7 t  
 ORIGIN

Query Match 66.0%; Score 13.2; DB 9; Length 33;  
 Best Local Similarity 83.3%; Pred. No. 3.3e+04;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 tgatgcttaaaagcttac 20  
 | |||| ||||| ||  
 Db 28 TTATGCTTAAAGCTTC 11

RESULT 9  
 I92840/c  
 LOCUS I92840 33 bp DNA 01-DEC-1998  
 DEFINITION Sequence 14 from patent US 5728821.  
 ACCESSION I92840 PAT  
 VERSION I92840.1 GI:3937310  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 33)  
 AUTHORS Yelton, D., Glaser, S., Huse, W. and Rosok, M. Joanne.  
 TITLE Mutant BR96 antibodies reactive with human carcinomas

JOURNAL Patent: US 5728821-A 14 17-MAR-1998;  
 FEATURES Location/Qualifiers  
 source  
 1. .33  
 /organism="unknown"  
 BASE COUNT 12 a 7 c 7 g 7 t  
 ORIGIN

Query Match 66.0%; Score 13.2; DB 10; Length 33;  
 Best Local Similarity 83.3%; Pred. No. 3.3e+04;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 tgatgcttaaaagcttac 20  
 | |||| ||||| ||  
 Db 28 TTATGCTTAAAGCTTC 11

RESULT 10  
 A61915/c  
 LOCUS A61915 35 bp DNA 09-MAR-1998  
 DEFINITION Sequence 3 from Patent WO9712969.  
 ACCESSION A61915 PAT  
 VERSION A61915.1 GI:3716021  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 35)  
 AUTHORS Poustka, A. and Coy, J.  
 TITLE FMRI-RELATED PROTEIN  
 JOURNAL Patent: WO 9712969-A 3 10-APR-1997;  
 DEUTSCHES KREBSFORSCH (DE)  
 COMMENT Other publication DE 19534763 970515.  
 FEATURES  
 source  
 1. .35  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 10 a 7 c 7 g 11 t  
 ORIGIN

Query Match 63.0%; Score 12.6; DB 9; Length 35;  
 Best Local Similarity 78.9%; Pred. No. 6.4e+04;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 atgatgcttaaaagcttac 20  
 || || ||||| ||||| ||  
 Db 20 ATAAATCTTAAAGCTTCC 2

RESULT 11  
 AR049871/c  
 LOCUS AR049871 25 bp DNA 29-SEP-1999  
 DEFINITION Sequence 10 from patent US 5824777.  
 ACCESSION AR049871 PAT  
 VERSION AR049871.1 GI:5971863  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 25)  
 AUTHORS Sasaki, K., Mori, T. and Makino, S.  
 TITLE Attenuated measles virus vaccine containing specific nucleotide  
 sequence and a method for its absolute identification  
 JOURNAL Patent: US 5824777-A 10 20-OCT-1998;  
 FEATURES  
 source  
 1. .25  
 Location/Qualifiers  
 BASE COUNT 8 a 7 c 6 g 4 t  
 ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 25;

Best Local Similarity 82.4%; Pred. No. 1e+05; Mismatches 0; Indels 3; Gaps 0;

QY 1 catgatgcttaaaagct 17  
||||| | | | |

Db 25 CATGATGCTTGAGAGGT 9

RESULT 12  
I59435/c

LOCUS I59435 25 bp DNA PAT 07-OCT-1997

DEFINITION Sequence 10 from patent US 5654136.

ACCESSION I59435

VERSION I59435.1 GI:2478067

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 25)  
AUTHORS Sasaki,K., Mori,T. and Makino,S.  
TITLE Attenuated measles virus vaccine, containing specific nucleotide sequence and a method for its absolute identification

JOURNAL Patent: US 5654136-A 10 05-AUG-1997;

FEATURES  
Location/Qualifiers  
1..25  
/organism="unknown"

BASE COUNT 8 a 7 c 6 g 4 t

ORIGIN

Query Match 61.0%; Score 12.2; DB 10; Length 25;  
Best Local Similarity 82.4%; Pred. No. 1e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 catgatgcttaaaagct 17  
-||||| | | | |

Db 25 CATGATGCTTGAGAGGT 9

RESULT 13  
AR110319/c

LOCUS AR110319 27 bp DNA PAT 14-FEB-2001

DEFINITION Sequence 13 from patent US 6114503.

ACCESSION AR110319

VERSION AR110319.1 GI:12826595

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 27)  
AUTHORS Wei,Y., Ruben,S.M., Sancar,A., Hsu,S.D. and Kazantsev,A.G.  
TITLE Human blue-light photoreceptor hCRY2

JOURNAL Patent: US 6114503-A 13 05-SEP-2000;

FEATURES  
Location/Qualifiers  
1..27  
/organism="unknown"

BASE COUNT 4 a 8 c 8 g 7 t

ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 27;  
Best Local Similarity 82.4%; Pred. No. 1e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 atgatgcttaaaagctt 18  
| | | | | | | | | |

Db 20 AGGATGCTGAAAGCTT 4

RESULT 14  
A65301

LOCUS A65301 29 bp DNA PAT 29-MAR-1999

DEFINITION Sequence 24 from patent WO9735983.

ACCESSION A65301.1 GI:4531080

KEYWORDS

SOURCE unidentified.

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 29)  
AUTHORS Jepson,I., Greenland,A.J., Thomas,D.R. and Philippe.  
TITLE CYSTEINE PROTEASE PROMOTER FROM OIL SEED RAPE AND A METHOD FOR THE CONTAINMENT OF PLANT GERMPLASM

JOURNAL Patent: WO 9735983-A 24 02-OCT-1997;  
ZENECA LTD (GB)

COMMENT Other publication AU 2033797 19971017.

FEATURES  
Location/Qualifiers  
1..29  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/clone="CYSGI2CR"

BASE COUNT 7 a 4 c 8 g 10 t

ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 29;  
Best Local Similarity 82.4%; Pred. No. 1e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 atgatgcttaaaagctt 18  
| | | | | | | | | |

Db 2 ATGATGGTGACAGCTT 18

RESULT 15  
AR036134/c

LOCUS AR036134 29 bp DNA PAT 29-SEP-1999

DEFINITION Sequence 4 from patent US 5871991.

ACCESSION AR036134

VERSION AR036134.1 GI:5952802

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 29)  
AUTHORS Elrod,S.L. and Cherry,J.R.  
TITLE Aspergillus oryzae 5-aminolevulinic acid synthases and nucleic acids encoding same

JOURNAL Patent: US 5871991-A 4 16-FEB-1999;

FEATURES  
Location/Qualifiers  
1..29  
/organism="unknown"

BASE COUNT 7 a 8 c 8 g 6 t

ORIGIN

Query Match 61.0%; Score 12.2; DB 9; Length 29;  
Best Local Similarity 82.4%; Pred. No. 1e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tgatgcttaaaagctta 19  
| | | | | | | | | |

Db 18 TGCTGCTTAAATGCATA 2

Search completed: October 2, 2001, 15:56:52  
Job time: 14176 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:08 ; Search time 10798.2 Seconds  
(without alignments)  
17.508 Million cell updates/sec

Title: US-09-757-100B-20

Perfect score: 20

Sequence: 1 catgatgcttaaaagcttac 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				TITLE	
Result No.	Score	Query Match	Length DB ID	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts	
1	13.2	66.0	50 243	Unpublished (2000)	
2	13	65.0	27 249	Contact: Robert B. Weiss	
3	13	65.0	39 189	University of Utah	
4	12.8	64.0	37 258	Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA	
5	12.2	61.0	30 250	Tel: 801 585 5606	
6	12.2	61.0	31 258	Fax: 801 585 7177	
7	12.2	61.0	37 15	Email: ddunn@genetics.utah.edu	
8	12.2	61.0	48 141	Insert Length: 10000 Std Error: 0.00	
9	12.2	61.0	48 141	Plate: 0171 row: I column: 17	
10	11.8	59.0	47 258	Seq primer: CGTTGTAAACGACGGCCAGT	
11	11.8	59.0	48 258	Class: plasmid ends	
12	11.8	59.0	49 258	High quality sequence stop: 50.	
13	11.6	58.0	32 258	Location/Qualifiers	
14	11.6	58.0	38 244	1. 50	
15	11.6	58.0	47 258	/organism="Mus musculus"	
16	11.6	58.0	50 107	/strain="C57BL/6J"	
17	11.6	58.0	50 107	/db_xref="taxon:10090"	
18	11.4	57.0	36 249	/clone="UUGC1M017117"	
19	11.4	57.0	40 242	/clone_lib="Mouse 10kb plasmid UUGC1M library"	
20	11.2	56.0	34 250	/sex="Male"	
21	11.2	56.0	34 250	/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"	
22	11.2	56.0	40 8	/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource	
23	11.2	56.0	44 247	(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114[gb]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."	
24	11.2	56.0	45 244	BASE COUNT 12 a 9 c 13 g 16 t	
25	11.2	56.0	47 152	ORIGIN	
26	11.2	56.0	49 6	Query Match 66.0%; Score 13.2; DB 243; Length 50;	
27	11	55.0	34 244	Best Local Similarity 83.3%; Pred. No. 1.7e+04;	
28	11	55.0	38 250	Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
29	11	55.0	41 244	QY 3 tgatgcttaaaagcttac 20	
30	11	55.0	46 243		
31	11	55.0	46 249	Db 11 TTATGCTTAAATCTCAC 28	
32	11	55.0	46 258	RESULT 2	
33	11	55.0	50 107	AZ785639 27 bp DNA GSS 16-FEB-2001	
34	11	55.0	50 156	LOCUS 2M0029012R Mouse 10kb plasmid UUGC1M library Mus musculus genomic	
35	10.8	54.0	21 246	DEFINITION clone UUGC2M0029012 R, DNA sequence.	
36	10.8	54.0	35 249	ACCESSION AZ785639	
37	10.8	54.0	36 249	VERSION AZ785639.1	
38	10.8	54.0	43 14	KEYWORDS GSS.	
39	10.8	54.0	43 14	SOURCE house mouse.	
40	10.8	54.0	43 241	ORGANISM Mus musculus	
41	10.8	54.0	43 241	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
42	10.8	54.0	49 244	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
43	10.8	54.0	50 107	1 (bases 1 to 50)	
44	10.6	53.0	28 249	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.	
45	10.6	53.0	32 249	and Wright,D.,Weiss,R.	

TITLE and Wright,D., Weiss,R.  
 JOURNAL Mouse whole genome scaffolding with paired end reads from 10kb  
 COMMENT plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0029 row: 0 column: 12  
 Seq primer: CACACAGGAAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 27.  
 Location/Qualifiers  
 1. 27  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUC240029012"  
 /clone\_lib="Mouse 10Kb plasmid UUC1m library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gii4732114|gii4732114|gii4732114), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## FEATURES

source

1. 27

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUC240029012"

/clone\_lib="Mouse 10Kb plasmid UUC1m library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gii4732114|gii4732114|gii4732114), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 17 a 3 c 2 g 5 t

ORIGIN

Query Match 65.0%; Score 13; DB 249; Length 27;

Best Local Similarity 100.0%; Pred. No. 2e+04;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 atgatgctataaa 14

|||||

Db 5 ATGATGCTTAAAA 17

RESULT 3

LOCUS U74031/c 39 bp mRNA EST 05-SEP-1997

DEFINITION RNU74031 rat lambda ZAPII library (C.P.Hamel) Rattus norvegicus

cDNA clone pc083, mRNA sequence.

ACCESSION U74031

VERSION U74031.1 GI:1658134

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 39)

AUTHORS Soto-Prior,A., Lavigne-Rebillard,M., Lenoir,M., Ripoll,C.,

TITLE Rebillard,G., Vago,P., Pujol,R. and Hamel,C.P.  
 JOURNAL Identification of preferentially expressed cochlear genes by  
 COMMENT systematic sequencing of a rat cochlea cDNA library  
 Brain Res. Mol. Brain Res. 47 (1-2), 1-10 (1997)  
 97364947  
 Contact: Hamel, CP  
 Laboratoire de Neurobiologie de l'Audition - INSERM U254  
 Institut National de la Sante et de la Recherche Medicale  
 C.H.R. St-Charles, 34295 Montpellier cedex, France  
 Tel: (33) 04 67 33 69 75  
 Fax: (33) 04 67 52 56 01  
 Email: biomol@net.fr  
 also expressed in brain, cerebellum, eye, lung, kidney and liver as  
 determined by RT-PCR.

## FEATURES

source

1. 39

/organism="Rattus norvegicus"

/strain="Wistar"

/db\_xref="taxon:10116"

/clone="pc083"

/clone\_lib="rat lambda ZAPII library (C.P.Hamel)"

/dev\_stage="postnatal day 24"

/lab\_host="Escherichia coli"

/note="Organ: cochlea; Vector: lambda ZAPII; Site\_1: EcoRI

; Site\_2: XhoI"

BASE COUNT 15 a 9 c 4 g 11 t

ORIGIN

Query Match 65.0%; Score 13; DB 189; Length 39;

Best Local Similarity 100.0%; Pred. No. 2.1e+04;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 tgcttaaaagctt 18

|||||

Db 14 TGCCTAAAGCTT 2

RESULT 4

LOCUS TA37H11P 37 bp DNA GSS 13-DEC-2000

DEFINITION T. brucei sheared genomic DNA clone 37h11, forward sequence,

genomic survey sequence.

ACCESSION AL453069

VERSION AL453069.1 GI:11854580

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE 1 (bases 1 to 37)

AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,

Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,

Melville,S.E., Rajandream,M.A. and Barrell,B.G.

Direct Submission

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,

Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

nh@sanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of

trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared

to give a tight size distribution (

4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small

insert libraries for whole genome shotgun sequencing projects. In

Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available

at http://www.sanger.ac.uk/Projects/T\_brucei/.

Location/Qualifiers

1. 37

FEATURES

source

/organism="Trypanosoma brucei"  
/strain="TREU927"  
/db\_xref="taxon:5691"  
/clone="37h11"

BASE COUNT 8 a 6 c 6 g 17 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 258; Length 37;  
Best Local Similarity 87.5%; Pred. No. 2.7e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 tgatgcttaaaagctt 18  
||||| || |||||

Db 3 TGATGCTTGAATGCTT 18

## RESULT 5

AZ833603 30 bp DNA GSS 20-FEB-2001  
LOCUS  
DEFINITION 2M011516R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M011516 R, DNA sequence.

ACCESSION AZ833603

VERSION AZ833603.1 GI:13003511

KEYWORDS GSS.

SOURCE house musculus

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 30)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

CONTACT: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0115 row: J column: 16

Seq primer: CACACGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 30.

Location/Qualifiers

## FEATURES

source

1..30

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0115J16"

/clone.lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (g14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 11 a 6 c 4 g 9 t

## ORIGIN

Query Match 61.0%; Score 12.2; DB 250; Length 30;  
Best Local Similarity 82.4%; Pred. No. 5.2e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 catgatgcttaaaagct 17

||||| || ||||| |||||

Db 5 CCTTATACTTTAAAGCT 21

## RESULT 6

TA52H11Q

LOCUS

DEFINITION TA52H11Q 31 bp DNA GSS 13-DEC-2000

T. brucei sheared genomic DNA clone 52h11, reverse sequence,

genomic survey sequence.

ACCESSION AL456059

VERSION AL456059.1 GI:11856930

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE 1 (bases 1 to 31)

AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,

Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,

Meiville,S.E., Rajandream,M.A. and Barrell,B.G.

Direct Submission

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,

Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

nh@sanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared

to give a tight size distribution (

4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small

insert libraries for whole genome shotgun sequencing projects. In

Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available

at http://www.sanger.ac.uk/Projects/T\_brucei/.

Location/Qualifiers

source

1..31

/organism="Trypanosoma brucei"

/strain="TREU927"

/db\_xref="taxon:5691"

/clone="52h11"

BASE COUNT 9 a 5 c 6 g 11 t

ORIGIN

Query Match 61.0%; Score 12.2; DB 258; Length 31;  
Best Local Similarity 82.4%; Pred. No. 5.2e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 tgatgcttaaaagctta 19

||||| ||||| |||||

Db 6 TCAGGGTTAATAGCTTA 22

## RESULT 7

AI024143

LOCUS

AI024143

37 bp mRNA EST 27-AUG-1998

```

DEFINITION  ov73a08.sl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1642934
              3' similar to gb:D13748 EUKARYOTIC INITIATION FACTOR 4A-I (HUMAN);,
              mRNA sequence.
ACCESSION   AI024143
VERSION     AI024143.1  GI:3239187
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 37)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE       Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/dbp/image/image.html

Trace considered overall poor quality
Insert Length: 669 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers
    1..37
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:1642934"
    /clone_lib="Soares_testis_NHT"
    /sex="male"
    /lab_host="DH10B"
    /note="Vector: p7T73D-Pac (Pharmacia) with a modified
    polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
    was prepared from mRNA obtained from Clontech Laboratories
    , Inc., and primed with a Not I - oligo(dT) primer [5'.
    TGTTACCAATCGAAGTGGAGCGCCGCCAATTTTTTTTTTTT 3'].
    Double-stranded cDNA was ligated to Eco RI adaptors
    (Pharmacia), digested with Not I and cloned into the Not I
    and Eco RI sites of the modified pT73 vector. Library
    went through one round of normalization to Cot5 and was
    constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT  12 a  b c  9 g  8 t
ORIGIN
Query Match      61.0%; Score 12.2; DB 15; Length 37;
Best Local Similarity 82.4%; Pred No. 5.3e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 catgatgtctaaagact 17
    ||||| |||| || ||
Db 19 CATGATACTTAGAAACT 35

RESULT 8
BF144195 38 bp mRNA EST 24-OCT-2000
LOCUS 601786804F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4014744 5',
DEFINITION mRNA sequence.
ACCESSION BF144195
VERSION BF144195
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 38)

```

```

AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9259 row: g column: 01
            High quality sequence stop: 38.
            Location/Qualifiers
    1..38
    /organism="Mus musculus"
    /strain="CZECH II"
    /db_xref="taxon:10090"
    /clone="IMAGE:4014744"
    /clone_lib="NCI_CGAP_Lu30"
    /tissue_type="tumor, metastatic to mammary"
    /lab_host="DH10B"
    /note="Organ: lung; Vector: pCMV-SPORT6; Site.1: NotI;
    Site.2: SalI; transgenic model WNT-1, expression driven by
    MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
    dT. Library constructed by Life Technologies.
    Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT  9 a  9 c  11 g  9 t
ORIGIN
Query Match      61.0%; Score 12.2; DB 145; Length 38;
Best Local Similarity 82.4%; Pred No. 5.3e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gatgctaaaagcttac 20
    ||||| ||||| |
Db 18 GATGCATTAAAGCTTGC 34

RESULT 9
BE896146 48 bp mRNA EST 20-OCT-2000
LOCUS 601438738F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3923841 5',
DEFINITION mRNA sequence.
ACCESSION BE896146
VERSION BE896146.1  GI:10360075
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 48)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTF
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9760 row: k column: 10
            High quality sequence stop: 48.
            Location/Qualifiers
    1..48
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:3923841"

```



Db 20 GATGCTTAACAGCCT 34

RESULT 12  
 TA213B010/c  
 LOCUS TA213B010 49 bp DNA GSS 13-DEC-2000  
 DEFINITION T. brucei sheared genomic DNA clone 213B01, reverse sequence, genomic survey sequence.  
 ACCESSION AL478475  
 VERSION AL478475.1 GI:11844434  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 REFERENCE 1 (bases 1 to 49)  
 AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrellesanger.ac.uk and nhlesanger.ac.uk  
 COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).  
 Email: nelsayedetigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).  
 FEATURES  
 source  
 1..49  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="213B01"  
 BASE COUNT 17 a 12 c 3 g 17 t  
 ORIGIN  
 17 a 12 c 3 g 17 t  
 Query Match 59.0%; Score 11.8; DB 258; Length 49;  
 Best Local Similarity 86.7%; Pred: No. 8.7e+04;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 atgatgcttaaaagc 16  
 Db 19 ATGATGATTAAAGGC 5  
 Query Match 59.0%; Score 11.8; DB 258; Length 49;  
 Best Local Similarity 86.7%; Pred: No. 8.7e+04;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 atgatgcttaaaagc 16  
 Db 19 ATGATGATTAAAGGC 5  
 RESULT 13  
 T73552  
 LOCUS T73552 50 bp mRNA EST 02-MAR-1995  
 DEFINITION Yc35h10.r1 Stragene liver (#937224) Homo sapiens cDNA clone IMAGE:82723 5' similar to gb:304449 CYTOCHROME P450 IIIA4 (HUMAN); mRNA sequence.  
 ACCESSION T73552  
 VERSION T73552.1 GI:690227  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 50)  
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissole, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,

Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996)  
 97044478  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Insert Size: 873  
 High quality sequence starts: 1 High quality sequence stops: 1  
 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Insert Length: 873 Std Error: 0.00  
 Seq primer: M13RP1  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1..50  
 /organism="Homo sapiens"  
 /db\_xref="GDB:499780"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:82723"  
 /clone.lib="Stratagene liver (#937224)"  
 /sex="male"  
 /dev\_stage="49 years old"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: liver; Vector: pBluescript SK; Site\_1: EcoRI ; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Hepatectomy from normal male caucasian. Average insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"  
 BASE COUNT 10 a 7 c 9 g 17 t  
 ORIGIN  
 10 a 7 c 9 g 17 t  
 Query Match 59.0%; Score 11.8; DB 189; Length 50;  
 Best Local Similarity 72.2%; Pred: No. 8.7e+04;  
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 catgatgcttaaaagcctt 18  
 Db 14 CANGAGCTTTNTAAGATT 31  
 RESULT 14  
 TA318D05P/c  
 LOCUS TA318D05P 32 bp DNA GSS 13-DEC-2000  
 DEFINITION T. brucei sheared genomic DNA clone 318D05, forward sequence, genomic survey sequence.  
 ACCESSION AL492652  
 VERSION AL492652.1 GI:11867479  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 REFERENCE 1 (bases 1 to 32)  
 AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrellesanger.ac.uk and nhlesanger.ac.uk  
 COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (



4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).  
Email: nelsayed@tigr.org  
Details of T. brucei sequencing at the Sanger Centre are available at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

## FEATURES

source  
1..32  
/organism="Trypanosoma brucei"  
/strain="PREU927"  
/db\_xref="taxon:5691"  
/clone="318d05"  
8 a 6 c 5 g 13 t

BASE COUNT  
ORIGIN

Query Match 58.0%; Score 11.6; DB 258; Length 32;  
Best Local Similarity 77.8%; Pred. No. 1e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 atgatgtcttaaaagctta 19  
||||| |||||||||  
Db 26 ATGAAGTGTAAAGCTTA 9

## RESULT 15

AZ487122/c  
LOCUS AZ487122 38 bp DNA GSS 05-OCT-2000  
DEFINITION IM0316P22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0316P22 F, DNA sequence.  
ACCESSION AZ487122  
VERSION AZ487122.1 GI:10654557  
KEYWORDS GSS.  
SOURCE house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 38)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0316 row: p column: 22

Seq primer: CGTGTAAACGAGGCCAGT

Class: plasmid ends

High quality sequence stop: 38.

## FEATURES

source  
1..38  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0316P22"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: pWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 13 a 4 c 9 g 12 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 244; Length 38;  
Best Local Similarity 77.8%; Pred. No. 1.1e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 tgatgcttaaaagcttac 20  
| |||| |||| ||||  
Db 30 TAATGCATAAATGCTCAC 13

Search completed: October 2, 2001, 15:01:10  
Job time: 10839 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:52 ; Search time 3339.34 Seconds  
(without alignments)  
69.480 Million cell updates/sec

Title: US-09-757-100B-23  
Perfect score: 15  
Sequence: 1 gcgggctcacagtgg 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*

1:	gb_ba1:*
2:	gb_ba2:*
3:	gb_ba3:*
4:	gb_in1:*
5:	gb_in2:*
6:	gb_in3:*
7:	gb_cm:*
8:	gb_ov:*
9:	gb_pat1:*
10:	gb_pat2:*
11:	gb_ph:*
12:	gb_pl1:*
13:	gb_pl2:*
14:	gb_pl3:*
15:	gb_pl4:*
16:	em_ba1:*
17:	em_ba2:*
18:	em_fun:*
19:	em_htgo_hum:*
20:	em_htgo_inv:*
21:	em_htgo_rod:*
22:	em_htg_hum1:*
23:	em_htg_hum2:*
24:	em_htg_hum3:*
25:	em_htg_hum4:*
26:	em_htg_hum5:*
27:	em_htg_hum6:*
28:	em_htg_hum7:*
29:	em_htg_hum8:*
30:	em_htg_inv1:*
31:	em_htg_inv2:*
32:	em_htg_other:*
33:	em_htg_rod:*
34:	em_hum1:*
35:	em_hum2:*
36:	em_hum3:*
37:	em_hum4:*
38:	em_hum5:*
39:	em_hum6:*
40:	em_hum7:*
41:	em_in:*
42:	em_cm:*
43:	em_or:*

44:	em_ov:*
45:	em_pat:*
46:	em_ph:*
47:	em_pl:*
48:	em_ro:*
49:	em_sts:*
50:	em_sy:*
51:	em_un:*
52:	em_vi:*
53:	gb_sts1:*
54:	gb_sts2:*
55:	gb_sts3:*
56:	gb_sy:*
57:	gb_un:*
58:	gb_vil:*
59:	gb_vl2:*
60:	gb_htg1:*
61:	gb_htg2:*
62:	gb_htg3:*
63:	gb_htg4:*
64:	gb_htg5:*
65:	gb_htg6:*
66:	gb_htg7:*
67:	gb_htg8:*
68:	gb_htg9:*
69:	gb_htg10:*
70:	gb_htg11:*
71:	gb_htg12:*
72:	gb_htg13:*
73:	gb_htg14:*
74:	gb_htg15:*
75:	gb_htg16:*
76:	gb_htg17:*
77:	gb_htg18:*
78:	gb_htg19:*
79:	gb_htg20:*
80:	gb_htg21:*
81:	gb_htg22:*
82:	gb_htg23:*
83:	gb_htg24:*
84:	gb_htg25:*
85:	gb_pr1:*
86:	gb_pr2:*
87:	gb_pr3:*
88:	gb_pr4:*
89:	gb_pr5:*
90:	gb_pr6:*
91:	gb_pr7:*
92:	gb_pr8:*
93:	gb_pr9:*
94:	gb_rol:*
95:	gb_ro2:*
96:	gb_in4:*
97:	gb_pr10:*
98:	em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	11.4	76.0	18	9 AR078873	AR078873 Sequence
c 2	11.4	76.0	22	10 I76917	I76917 Sequence 25
c 3	11	73.3	21	10 E26933	E26933 Vascular en
c 4	11	73.3	25	9 AR090736	AR090736 Sequence
c 5	11	73.3	31	10 I11623	I11623 Sequence 8
c 6	11	73.3	37	95 S45430	S45430 V alpha J a
7	10.8	72.0	18	9 AR080704	AR080704 Sequence
8	10.8	72.0	24	9 AR019663	AR019663 Sequence

RESULT 1  
AR078873/c  
LOCUS AR078873 18 bp DNA PAT 31-AUG-2000  
DEFINITION Sequence 17 from patent US 5965370.  
ACCESSION AR078873  
VERSION AR078873.1 GI:10005619  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 18)  
AUTHORS Cowser, L.M.  
TITLE Antisense modulation of RhoG expression  
JOURNAL Patent: US 5965370-A 17 12-OCT-1999;  
FEATURES  
source 1..18  
BASE COUNT 1 a 6 c 6 g 5 t  
ORIGIN

Query Match 76.0%; Score 11.4; DB 9; Length 18;  
Best Local Similarity 92.3%; Pred. No. 4.5e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cgggctcacagt 14  
|||||  
Db 17 CGGGCGCACAGT 5

RESULT 2  
AR070473 Sequence  
AR099765 Sequence  
I47711 Sequence 10  
AR082206 Sequence 50  
I78252 Sequence 50  
AX092192 Sequence  
AX007130 Sequence  
E51076 Novel Esche  
AR095613 Sequence  
AR095615 Sequence  
AR049882 Sequence  
AR098915 Sequence  
I79755 Sequence 51  
A60222 Sequence 18  
AR090849 Sequence 18  
I06514 Sequence 18  
X70713 M.musculus  
I46492 Sequence 47  
AR073339 Sequence  
A69911 Sequence 6  
AX041988 Sequence  
AR017555 Sequence  
AR084745 Sequence  
I21733 Sequence 8  
A58772 Sequence 18  
AX019995 Sequence  
AX020004 Sequence  
I19529 Sequence 6  
I19530 Sequence 7  
I19532 Sequence 9  
I86209 Sequence 6  
I86210 Sequence 7  
I86212 Sequence 9  
A58770 Sequence 16  
AR017553 Sequence 11  
I19534 Sequence 11  
I21731 Sequence 6

## ALIGNMENTS

RESULT 3  
E26933/c  
LOCUS E26933 21 bp DNA PAT 07-FEB-2001  
DEFINITION Vascular endothelial cell growth factor.  
ACCESSION E26933  
VERSION E26933.1 GI:13026353  
KEYWORDS JP 1999169183-A/13.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Hiroshi, M.S.C.C., Takashi, S.R.W.W. and Suzuki.  
TITLE Vascular endothelial cell growth factor  
JOURNAL Patent: JP 1999169183-A 13 29-JUN-1999;  
COMMENT AGENCY OF IND SCIENCE & TECHNOL, TOGGOSEI CHEM IND CO LTD  
OS Unidentified  
PN JP 1999169183-A/13  
PD 29-JUN-1999  
PF 11-DEC-1999 JP 1997362118  
PR  
PI HIROSHI MITSUI, SUNIL C KAURU, TAKASHI SUGIHARA, RENU WADAWA, PI  
HIDEO SUZUKI  
PC C12N15/09, A61K38/22, A61K38/22, A61K38/22, A61K38/22, A61K38/22,  
PC C07K14/52,  
PC C12N1/21, C12N5/10, C12P21/02, C12N1/21, C12R1:19, (C12N5/10, PC  
C12R1:91),  
PC (C12P21/02, C12R1:19), (C12P21/02, C12R1:91), C12N15/00, A61K37/24,  
PC A61K37/24,  
PC A61K37/24, A61K37/24, A61K37/24, C12N5/00, (C12N5/00, C12R1:91) CC  
Strandedness: Single;  
CC Topology: Linear;  
FH Key 1..21 Location/Qualifiers  
FT source 1..21  
Location/Qualifiers  
1..21  
/organism="unidentified"  
/db\_xref="taxon:32644"

FEATURES  
source  
BASE COUNT 4 a 5 c 6 g 6 t  
ORIGIN

Query Match 73.3%; Score 11; DB 10; Length 21;

```

Best Local Similarity 100.0%; Pred. No. 7.8e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ggctcacagt 14
Db 11 GGCTCACAGT 1

RESULT 4
AR090736/c 25 bp DNA PAT 07-SEP-2000
LOCUS AR090736
DEFINITION Sequence 856 from patent US 5994076.
ACCESSION AR090736
VERSION AR090736.1 GI:10017491
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 25)
AUTHORS Chenchik,A., Jokhadze,G. and Bibilashvilli,R.
TITLE Methods of assaying differential expression
JOURNAL Patent: US 5994076-A 856 30-NOV-1999;
FEATURES
    source
        1. .25
        /organism="unknown"
        6 a 9 c 6 g 4 t
BASE COUNT
ORIGIN

Query Match 73.3%; Score 11; DB 9; Length 25;
Best Local Similarity 100.0%; Pred. No. 7.5e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ggctcacagt 14
Db 11 GGCTCACAGT 1

RESULT 5
I11623/c 31 bp DNA PAT 26-JUL-1995
LOCUS I11623
DEFINITION Sequence 8 from Patent US 5408038.
ACCESSION I11623
VERSION I11623.1 GI:909141
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 31)
AUTHORS Smith,R.S., Curtiss,L.K., Koduri,K.R., Witztum,J.L. and Young,S.G.
TITLE Nonnatural apolipoprotein B-100 peptides and apolipoprotein
B-100-apolipoprotein A-I fusion peptides
JOURNAL Patent: US 5408038-A 8 18-APR-1995;
FEATURES
    source
        1. .31
        /organism="unknown"
        10 a 8 c 7 g 6 t
BASE COUNT
ORIGIN

Query Match 73.3%; Score 11; DB 10; Length 31;
Best Local Similarity 100.0%; Pred. No. 7.3e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gctcacagt 15
Db 19 GCTCACAGT 9

RESULT 6
S45430/c 37 bp DNA ROD 31-MAR-1995
LOCUS S45430
DEFINITION V alpha J alpha -T-cell receptor alpha chain [mice, BALB C, adult

day-37 thymus, Genomic, 37 nt].
S45430
VERSION S45430.1 GI:1679985
KEYWORDS
SOURCE Mus sp. BALB C adult day-37 thymus.
ORGANISM Mus sp.
REFERENCE 1 (bases 1 to 37)
AUTHORS Roth,M.E., Holman,P.O. and Kranz,D.M.
TITLE Nonrandom use of J alpha gene segments. Influence of V alpha and J
alpha gene location
JOURNAL J. Immunol. 147 (3), 1075-1081 (1991)
MEDLINE 91318150
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 45430] from the original journal article.
This sequence comes from Figure 2 B.
COMMENT On Nov 21, 1996 this sequence version replaced gi:1619731.
FEATURES
    source
        1. .37
        Location/Qualifiers
        /organism="Mus sp."
        /db_xref="taxon:10095"
        1. .37
        /partial
        /gene="VJaag;JJaag;"
        /note="T-cell receptor alpha chain"
        6 a 10 c 9 g 12 t
BASE COUNT
ORIGIN

Query Match 73.3%; Score 11; DB 95; Length 37;
Best Local Similarity 100.0%; Pred. No. 7.1e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgggctcacag 12
Db 21 CGGGCTCACAG 11

RESULT 7
AR080704 18 bp DNA PAT 31-AUG-2000
LOCUS AR080704
DEFINITION Sequence 9 from patent US 5968826.
ACCESSION AR080704
VERSION AR080704.1 GI:10007434
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 18)
AUTHORS Bennett,C.Frank, Condon,T.P. and Cowser,L.M.
TITLE Antisense inhibition of integrin.alpha.4 expression
JOURNAL Patent: US 5968826-A 9 19-OCT-1999;
FEATURES
    source
        1. .18
        Location/Qualifiers
        /organism="unknown"
        2 a 7 c 5 g 4 t
BASE COUNT
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 18;
Best Local Similarity 85.7%; Pred. No. 1.1e+05;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcgggctcacagt 14
Db 1 GCGGCTCTCAGT 14

RESULT 8
AR019663 24 bp DNA PAT 05-DEC-1998
LOCUS AR019663
DEFINITION Sequence 2 from patent US 5783683.

```

```
ACCESSION AR019663
VERSION AR019663.1 GI:3974777
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Morrison,R.S.
TITLE Antisense oligonucleotides which reduce expression of the FGFRI
gene
JOURNAL Patent: US 5783683-A 2 21-JUL-1998;
FEATURES Location/Qualifiers
source 1..24
BASE COUNT 5 a 6 c 7 g 6 t
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 24;
Best Local Similarity 85.7%; Pred. NO. 1e+05;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 15
||| ||||| |||||
Db 1 CGAGCTCACTGTGG 14

RESULT 9
AR070473/c AR070473 24 bp DNA PAT 18-FEB-2000
LOCUS AR070473
DEFINITION Sequence 9 from patent US 5905185.
ACCESSION AR070473
VERSION AR070473.1 GI:7221361
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Garner,I., Cottingham,I.R., Temperley,S.M., Foster,D.C.,
Sprecher,C.A. and Prunkard,D.E.
TITLE Protein C production in non-human transgenic mammals
JOURNAL Patent: US 5905185-A 9 18-MAY-1999;
FEATURES Location/Qualifiers
source 1..24
BASE COUNT 5 a 8 c 6 g 5 t
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 24;
Best Local Similarity 85.7%; Pred. NO. 1e+05;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 15
||| ||||| |||||
Db 24 CTGGCTCACAGG 11

RESULT 10
AR099765/c AR099765 24 bp DNA PAT 14-FEB-2001
LOCUS AR099765
DEFINITION Sequence 9 from patent US 6077990.
ACCESSION AR099765
VERSION AR099765.1 GI:12809531
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Leung,W., Darrow,A. and Andrade-Gordon,P.
TITLE PAR2 modified transgenic mice
JOURNAL Patent: US 6077990-A 9 20-JUN-2000;
FEATURES Location/Qualifiers

source 1..24
BASE COUNT 5 a 6 c 7 g 6 t
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 29;
Best Local Similarity 85.7%; Pred. NO. 9.9e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcgggctcacagtgg 14
||| ||||| ||||| |||||
```

---

```
source 1..24
/organism="unknown"
BASE COUNT 4 a 6 c 8 g 6 t
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 24;
Best Local Similarity 85.7%; Pred. NO. 1e+05;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 15
||| ||||| |||||
Db 15 CCGCTAACAGTGG 2

RESULT 11
I47711/c I47711 24 bp DNA PAT 07-OCT-1997
LOCUS I47711
DEFINITION Sequence 10 from patent US 5639940.
ACCESSION I47711
VERSION I47711.1 GI:2471676
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Garner,I., Dalrymple,M.L., Prunkard,D.E. and Foster,D.C.
TITLE Production of fibrinogen in transgenic animals
JOURNAL Patent: US 5639940-A 10 17-JUN-1997;
FEATURES Location/Qualifiers
source 1..24
BASE COUNT 5 a 8 c 6 g 5 t
ORIGIN

Query Match 72.0%; Score 10.8; DB 10; Length 24;
Best Local Similarity 85.7%; Pred. NO. 1e+05;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 15
||| ||||| ||||| |||||
Db 24 CTGGCTCACAGG 11

RESULT 12
AR082206/c AR082206 29 bp DNA PAT 31-AUG-2000
LOCUS AR082206
DEFINITION Sequence 50 from patent US 5972704.
ACCESSION AR082206
VERSION AR082206.1 GI:10008932
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 29)
AUTHORS Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and
Thompson,J.D.
TITLE HIV nef targeted ribozymes
JOURNAL Patent: US 5972704-A 50 26-OCT-1999;
FEATURES Location/Qualifiers
source 1..29
BASE COUNT 7 a 7 c 9 g 6 t
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 29;
Best Local Similarity 85.7%; Pred. NO. 9.9e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcgggctcacagtgg 14
||| ||||| ||||| |||||
```

Db 15 GCAGGCTCACAGG 2

## RESULT 13

178252/c 178252 29 bp DNA PAT 03-APR-1998  
LOCUS Sequence 50 from patent US 5693535.  
DEFINITION 178252  
ACCESSION 178252  
VERSION 178252.1 GI:3014406  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 29)  
AUTHORS Draper,K.G., Chowira,B., McSwiggen,J., Stinchcomb,D.T. and Thompson,J.D.  
TITLE HIV targeted ribozymes  
JOURNAL Patent: US 5693535-A 50 02-DEC-1997;  
FEATURES Location/Qualifiers  
source 1..29  
BASE COUNT 7 a 7 c 9 g 6 t  
ORIGIN

Query Match 72.0%; Score 10.8; DB 10; Length 29;  
Best Local Similarity 85.7%; Pred. No. 9.9e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcgggctcacagt 14  
|| |||||

Db 15 GCAGGCTCACAGG 2

## RESULT 14

AX092192 30 bp DNA PAT 21-MAR-2001  
LOCUS Sequence 8 from Patent WO0116321.  
DEFINITION AX092192  
ACCESSION AX092192  
VERSION AX092192.1 GI:13444410  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Tandon,N., Sun,B., Nakamura,T. and Yamamoto,N.  
TITLE Platelet membrane glycoprotein vi (gpvi) dna and protein sequences,  
and uses thereof  
JOURNAL Patent: WO 0116321-A 8 08-MAR-2001;  
OTSUKA PHARMACEUTICAL CO., LTD. (JP)  
FEATURES Location/Qualifiers  
source 1..30  
BASE COUNT 6 a 11 c 10 g 3 t  
ORIGIN

Query Match 72.0%; Score 10.8; DB 10; Length 30;  
Best Local Similarity 85.7%; Pred. No. 9.8e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 15  
||| |||||

Db 12 CGGCTCAGAGTGG 25

## RESULT 15

AX007130 36 bp DNA PAT 06-SEP-2000  
LOCUS Sequence 16 from Patent EP0972838.  
DEFINITION AX007130  
ACCESSION

VERSION AX007130.1 GI:9995021  
KEYWORDS synthetic construct.  
SOURCE synthetic construct  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 36)  
AUTHORS Schantz,C. and Kopetzki,E.  
TITLE Escherichia coli host/vector system based on antibiotic-free  
selection by complementation of an auxotrophy  
JOURNAL Patent: EP 0972838-A 16 19-JAN-2000;  
ROCHE DIAGNOSTICS GMBH (DE)  
FEATURES Location/Qualifiers  
source 1..36  
BASE COUNT 14 a 7 c 12 g 3 t  
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 36;  
Best Local Similarity 85.7%; Pred. No. 9.5e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 15  
||||| |

Db 12 CGGCTCAAGTGG 25

Search completed: October 2, 2001, 15:56:53  
Job time: 14177 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:10 ; Search time 10798.2 Seconds  
(without alignments)  
13.131 Million cell updates/sec

Title: US-09-757-100B-23

Perfect score: 15

Sequence: 1 gcgggctcacagtgg 15

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



**TITLE** and Wright,D.,Weiss,R.  
**JOURNAL** Mouse whole genome scaffolding with paired end reads from 10kb  
**COMMENT** plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0117 row: I column: 21  
 Seq primer: CGTTGTAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 50.  
**FEATURES** Location/Qualifiers  
 source  
 1..50  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUCG240117121"  
 /clone\_lib="Mouse 10Kb plasmid UUCG1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

**BASE COUNT** 14 a 10 c 23 g 3 t  
**ORIGIN**

Query Match 78.7%; Score 11.8; DB 250; Length 50;  
 Best Local Similarity 86.7%; Pred. No. 2e+04;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gggggtcacagtgg 15  
 |||||  
 Db 28 GCGGGCTCACTCTGG 14

**RESULT** 3  
**LOCUS** AU106328 50 bp mRNA EST 05-APR-2001  
**DEFINITION** AU106328 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 KAT00883, mRNA sequence.  
**ACCESSION** AU106328  
**VERSION** AU106328.1 GI:13555849  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 50)  
**AUTHORS** Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo

**TITLE** K., Suyama,A. and Sugano,S.  
**JOURNAL** Fine Structural analysis of transcription start sites of human  
**COMMENT** mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
 Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: ysuzuki@ims.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  
**FEATURES** Location/Qualifiers  
 source  
 1..50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="KAT00883"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
**BASE COUNT** 6 a 12 c 22 g 10 t  
**ORIGIN**

Query Match 76.0%; Score 11.4; DB 107; Length 50;  
 Best Local Similarity 92.3%; Pred. No. 3.3e+04;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ggggtcacagtgg 15  
 |||||  
 Db 22 GGGCCACAGTGG 34

**RESULT** 4  
**LOCUS** AU102352/c 50 bp mRNA EST 05-APR-2001  
**DEFINITION** AU102352 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 HRC01304, mRNA sequence.  
**ACCESSION** AU102352  
**VERSION** AU102352.1 GI:13551872  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 50)  
**AUTHORS** Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.  
 Fine Structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
 Unpublished (2001)  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: ysuzuki@ims.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

**BASE COUNT** 7 a 16 c 20 g 7 t  
**ORIGIN**

Query Match 73.3%; Score 11; DB 107; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+04;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gggggtcacaca 11



University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0401 row: K column: 03  
Seq primer: CTTGTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 38.

## FEATURES

Location/Qualifiers  
1. .38  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCLM0401K03"  
/clone\_lib="Mouse 10Kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g147321141q1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
ORIGIN  
9 a 8 c 9 g 12 t

Query Match 72.0%; Score 10.8; DB 246; Length 38;  
Best Local Similarity 85.7%; Pred. No. 7.1e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cgggctcacagtgg 15  
| | | | | | | | | |  
DB 2 CTGGATCAGTGG 15

## RESULT 8

AA822940/c  
LOCUS AA822940 41 bp mRNA EST 17-FEB-1998  
DEFINITION vp30g03.r1 Barstead mouse proximal colon MPLRB6 Mus musculus cDNA  
clone IMAGE:1078228 5' similar to gb:U13705 Mus musculus domesticus  
C57BL/6J plasma glutathione (MOUSE);, mRNA sequence.  
ACCESSION AA822940  
VERSION AA822940.1 GI:2892808  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 41)  
REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)

## COMMENT

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:592524

Trace considered overall poor quality  
Seq primer: -28ml3 rev2 Et from Amersham  
High quality sequence stop: 1.

## FEATURES

Location/Qualifiers  
1. .41  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1078228"  
/clone\_lib="Barstead mouse proximal colon MPLRB6"  
/dev\_stages="7 day juvenile"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA  
was primed with a Not I - oligo(df) primer 15'  
TGTACGAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTT  
3'); double-stranded cDNA was ligated to Eco RI adaptors  
[AATTCGATCCTTG], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT73 vector.  
Library constructed by Bob Barstead."

BASE COUNT  
ORIGIN  
11 a 12 c 12 g 6 t

Query Match 72.0%; Score 10.8; DB 12; Length 41;  
Best Local Similarity 85.7%; Pred. No. 7.2e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cgggctcacagtgg 15  
| | | | | | | | | |  
DB 27 CGCGTCACAGTTG 14

## RESULT 9

TA233A12Q  
LOCUS TA233A12Q 44 bp DNA GSS 13-DEC-2000  
DEFINITION T. brucei sheared genomic DNA clone 233a12, reverse sequence,  
genomic survey sequence.  
ACCESSION AL481198  
VERSION AL481198.1 GI:11846892  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 44)  
REFERENCE Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,  
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,  
Melville,S.E., Rajandream,M.A. and Barrell,B.G.  
Direct Submission  
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
nh@sanger.ac.uk  
COMMENT Constructed at the Institute for Genomic Research (TIGR),  
Rockville, MD. Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUFat 10.1) was mechanically sheared  
to give a tight size distribution (4 kb). The v + i method used for the library construction is  
described in detail in Smith, H. and Venter, J.C. (Making small  
insert libraries for whole genome shotgun sequencing projects. In  
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999).

Email: nelsayedetigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available  
 at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

```

FEATURES
  source
    1. .44
      /organism="Trypanosoma brucei"
      /strain="TREU927"
      /db_xref="taxon:5691"
      /clone="233a12"
BASE COUNT      12 a   10 c   12 g   10 t
ORIGIN

Query Match      72.0%; Score 10.8; DB 258; Length 44;
Best Local Similarity 85.7%; Pred. No. 7.2e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcgggtcacagtgg 14
    ||||| ||||| |
Db 16 GCGGGATCACAGGG 29

RESULT 10
AA934014/c
LOCUS
DEFINITION
  AA934014 48 bp mRNA EST 27-APR-1998
  om58e11.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1551404 3'
  similar to gb:X01683 ALPHA-1-ANTITRYPSIN PRECURSOR (HUMAN);, mRNA
  sequence.
ACCESSION
  AA934014.1 GI:3090282
VERSION
  AA934014.1
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 48)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncilogap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
JOURNAL
  Unpublished (1997)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: M. Bento Soares, Ph.D.
  DNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Seq primer: -40m13 fwd. Et from Amersham.
FEATURES
  source
    1. .48
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:1551404"
      /clone_lib="NCI_CGAP_GC4"
      /tissue_type="pooled germ cell tumors"
      /lab_host="DH10B"
      /notes="Vector: pT73D-Pac (Pharmacia) with a modified
      polylinker; 1st strand cDNA was prepared from 3 pooled
      germ cell tumors, and was then primed with a Not I -
      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
      RI adaptors (Pharmacia), digested with Not I and cloned
      into the Not I and Eco RI sites of the modified pT73
      vector. Library is normalized. Library was constructed by
      Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      7 a   16 c   15 g   10 t
ORIGIN

Query Match      72.0%; Score 10.8; DB 13; Length 48;
Best Local Similarity 85.7%; Pred. No. 7.3e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cggggtcacagtgg 15
    ||||| ||||| |
Db 19 CGGGCGCCAGTGG 32

RESULT 12
TA140A04P/c
LOCUS
DEFINITION
  T. brucei sheared genomic DNA clone 140a04, forward sequence,
  genomic survey sequence.
ACCESSION
  AL466405.1 GI:11835760
VERSION
  AL466405.1
KEYWORDS
  GSS.
SOURCE
  Trypanosoma brucei.
  Trypanosoma brucei.
  Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
  Trypanosoma.
REFERENCE
  1 (bases 1 to 22)
  Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
  Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
  Melville, S.E., Rajandream, M.A. and Barrell, B.G.
  Direct Submission
  Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
  project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
  Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

```

```

Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cggggtcacagtgg 15
    ||||| ||||| |
Db 30 CGGGTCACAGAGG 17

RESULT 11
AU107932
LOCUS
DEFINITION
  AU107932 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
  ZRV62008, mRNA sequence.
ACCESSION
  AU107932
VERSION
  AU107932.1 GI:13557454
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 50)
  Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata
  H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo
  K., Suyama, A. and Sugano, S.
  Fine Structural analysis of transcription start sites of human
  mRNAs using full-length enriched and 5'-end enriched cDNA libraries
  Unpublished (2001)
  Contact: Yutaka Suzuki
  Department of Virology
  Institute of Medical Science, University of Tokyo
  4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
  Email: ysuzuki@ims.u-tokyo.ac.jp
  Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
  S. Construction and characterization of a full length-enriched and
  a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES
  source
    1. .50
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="ZRV62008"
      /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT      6 a   19 c   15 g   9 t   1 others
ORIGIN

Query Match      72.0%; Score 10.8; DB 107; Length 50;
Best Local Similarity 85.7%; Pred. No. 7.3e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cggggtcacagtgg 15
    ||||| ||||| |
Db 19 CGGGCGCCAGTGG 32

RESULT 12
TA140A04P/c
LOCUS
DEFINITION
  T. brucei sheared genomic DNA clone 140a04, forward sequence,
  genomic survey sequence.
ACCESSION
  AL466405.1 GI:11835760
VERSION
  AL466405.1
KEYWORDS
  GSS.
SOURCE
  Trypanosoma brucei.
  Trypanosoma brucei.
  Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
  Trypanosoma.
REFERENCE
  1 (bases 1 to 22)
  Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
  Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
  Melville, S.E., Rajandream, M.A. and Barrell, B.G.
  Direct Submission
  Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
  project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
  Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

```

COMMENT  
 nh@sanger.ac.uk  
 Constructed at the Institute for Genomic Research (TIGR),  
 Rockville, MD. Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
 to give a tight size distribution (4 kb). The v + i method used for the library construction is  
 described in detail in Smith, H. and Venter, J.C. (Making small  
 insert libraries for whole genome shotgun sequencing projects. In  
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Barrell, Oxford University Press, 1999).  
 Email: nelsayed@tigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available  
 at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

#### FEATURES

1. .22  
 Location/Qualifiers  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="140a04"

BASE COUNT 3 a 6 c 10 g 3 t  
 ORIGIN

Query Match 69.3%; Score 10.4; DB 258; Length 22;  
 Best Local Similarity 91.7%; Pred. No. 1.1e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 4 ggctcacagtgg 15  
 ||| |||||  
 Db 13 GCGGCACAGTGG 2

RESULT 13  
 AZ770438/c  
 LOCUS AZ770438 25 bp DNA GSS 16-FEB-2001  
 DEFINITION IM0572G03F Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0572G03 F, DNA sequence.

ACCESSION AZ770438  
 VERSION AZ770438.1 GI:12891623  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 25)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 , M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL  
 COMMENT Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0572 row: G column: 03  
 Seq primer: CGTTGTAACGACGCGCAGT  
 Class: plasmid ends  
 High quality sequence stop: 25.

#### FEATURES

1. .25  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0572G03"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from Mus  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (<http://www.jax.org/resources/documents/dnares/>). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (gil4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."  
 BASE COUNT 5 a 5 c 9 g 6 t  
 ORIGIN

Query Match 69.3%; Score 10.4; DB 249; Length 25;  
 Best Local Similarity 91.7%; Pred. No. 1.1e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggctcacagtgg 15  
 ||| |||||  
 Db 15 GTCTCACAGTGG 4

RESULT 14  
 AZ345566/c

LOCUS AZ345566 31 bp DNA GSS 29-SEP-2000  
 DEFINITION IM0080I14F Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0080I14 F, DNA sequence.

ACCESSION AZ345566  
 VERSION AZ345566.1 GI:10424803  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 31)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 , M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL  
 COMMENT Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0080 row: I column: 14  
 Seq primer: CGTTGTAACGACGCGCAGT  
 Class: plasmid ends  
 High quality sequence stop: 31.

#### FEATURES

1. .31  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0080I14"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"



```

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      6 a      9 c      6 g      10 t
ORIGIN

```

```

Query Match      69.3%; Score 10.4; DB 242; Length 31;
Best Local Similarity 91.7%; Pred. No. 1.2e+05;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 gggtccacagt 14
    ||| |||||
Db 20 GGGATCAGTGG 9

```

```

RESULT 15
AZ497161/c
LOCUS
DEFINITION
  1M0333G22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
  clone UUGC1M0333G22 R, DNA sequence.
ACCESSION
  AZ497161
VERSION
  AZ497161.1 GI:10673865
KEYWORDS
  GSS.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
  1 (bases 1 to 35)
  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
  Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
  ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
  and Wright,D., Weiss,R.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
JOURNAL
  Unpublished (2000)
COMMENT
  Contact: Robert B. Weiss
  University of Utah Genome Center
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0333 row: G column: 22
  Seq primer: CACACAGGAACACGTATGACC
  Class: plasmid ends
  High quality sequence stop: 35.
  Location/Qualifiers
    1..35
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUGC1M0333G22"

```

```

FEATURES
  source

```

```

/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      10 a      9 c      9 g      7 t
ORIGIN

```

```

Query Match      69.3%; Score 10.4; DB 245; Length 35;
Best Local Similarity 91.7%; Pred. No. 1.2e+05;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 gggtccacagtgg 15
    ||| |||||
Db 14 GTCTCAGTGG 3

```

```

Search completed: October 2, 2001, 15:01:11
Job time: 10840 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:53 ; Search time 3339.34 Seconds  
(without alignments)  
59.480 Million cell updates/sec

Title: US-09-757-100B-30  
Perfect score: 15  
Sequence: 1 taagcagctgccatt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_ba3:\*

4: gb\_in1:\*

5: gb\_in2:\*

6: gb\_in3:\*

7: gb\_om:\*

8: gb\_ov:\*

9: gb\_pat1:\*

10: gb\_pat2:\*

11: gb\_ph:\*

12: gb\_pl1:\*

13: gb\_pl2:\*

14: gb\_pl3:\*

15: gb\_pl4:\*

16: em\_ba1:\*

17: em\_ba2:\*

18: em\_fun:\*

19: em\_htgo\_hum:\*

20: em\_htgo\_inv:\*

21: em\_htgo\_rod:\*

22: em\_htg\_hum1:\*

23: em\_htg\_hum2:\*

24: em\_htg\_hum3:\*

25: em\_htg\_hum4:\*

26: em\_htg\_hum5:\*

27: em\_htg\_hum6:\*

28: em\_htg\_hum7:\*

29: em\_htg\_hum8:\*

30: em\_htg\_inv1:\*

31: em\_htg\_inv2:\*

32: em\_htg\_other:\*

33: em\_htg\_rod:\*

34: em\_hum1:\*

35: em\_hum2:\*

36: em\_hum3:\*

37: em\_hum4:\*

38: em\_hum5:\*

39: em\_hum6:\*

40: em\_hum7:\*

41: em\_in:\*

42: em\_om:\*

43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vi:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vil:\*

59: gb\_vil2:\*

60: gb\_vil2:\*

61: gb\_htg1:\*

62: gb\_htg2:\*

63: gb\_htg3:\*

64: gb\_htg4:\*

65: gb\_htg5:\*

66: gb\_htg6:\*

67: gb\_htg7:\*

68: gb\_htg8:\*

69: gb\_htg9:\*

70: gb\_htg10:\*

71: gb\_htg11:\*

72: gb\_htg12:\*

73: gb\_htg13:\*

74: gb\_htg14:\*

75: gb\_htg15:\*

76: gb\_htg16:\*

77: gb\_htg17:\*

78: gb\_htg18:\*

79: gb\_htg19:\*

80: gb\_htg20:\*

81: gb\_htg21:\*

82: gb\_htg22:\*

83: gb\_htg23:\*

84: gb\_htg24:\*

85: gb\_htg25:\*

86: gb\_pr1:\*

87: gb\_pr2:\*

88: gb\_pr3:\*

89: gb\_pr4:\*

90: gb\_pr5:\*

91: gb\_pr6:\*

92: gb\_pr7:\*

93: gb\_pr8:\*

94: gb\_pr9:\*

95: gb\_ro1:\*

96: gb\_ro2:\*

97: gb\_in4:\*

98: gb\_pri0:\*

99: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	13.4	89.3	18	10	I44719 Sequence 45
c 2	13.4	89.3	18	10	I52125 Sequence 45
c 3	11.8	78.7	50	9	AR032874 Sequence
c 4	11.8	78.7	50	9	AR032875 Sequence
c 5	11.8	78.7	50	10	I29614 Sequence 48
c 6	11.8	78.7	50	10	I29615 Sequence 48
c 7	11.8	78.7	50	10	I91288 Sequence 48
c 8	11.8	78.7	50	10	I91289 Sequence 48

```

c 9 11.4 76.0 21 10 AX094845
c 10 11.4 76.0 22 9 A80791
c 11 11.4 76.0 22 9 AX022755
c 12 11.4 76.0 25 9 AX016058
c 13 11.4 76.0 25 9 AR075530
c 14 11.4 76.0 25 9 AX019566
c 15 11.4 76.0 26 10 E04774
c 16 11.4 76.0 30 9 A84595
c 17 11.4 76.0 30 9 AR079892
c 18 11.4 76.0 39 9 AR032341
c 19 11.4 76.0 39 9 AR080603
c 20 11.4 76.0 41 9 A82505
c 21 11.4 76.0 41 9 A82526
c 22 11 73.3 20 45 E10367
c 23 11 73.3 22 9 AR030684
c 24 11 73.3 27 9 A83874
c 25 11 73.3 28 9 AR026168
c 26 11 73.3 28 9 AR026220
c 27 11 73.3 28 9 AR026234
c 28 11 73.3 28 9 AR026248
c 29 11 73.3 28 10 E61327
c 30 11 73.3 28 10 I82910
c 31 11 73.3 28 10 I82962
c 32 11 73.3 28 10 I82976
c 33 11 73.3 28 10 I82990
c 34 11 73.3 29 10 I46944
c 35 11 73.3 30 9 A12174
c 36 11 73.3 32 10 I06269
c 37 11 73.3 35 10 I74286
c 38 11 73.3 39 9 A28773
c 39 11 73.3 39 9 A57638
c 40 11 73.3 39 9 AR088301
c 41 11 73.3 41 9 AR032336
c 42 11 73.3 45 9 AX048222
c 43 11 73.3 45 9 AX049383
c 44 11 73.3 45 9 AX052723
c 45 11 73.3 45 9 AX053164

ALIGNMENTS

RESULT 1
144719/c
LOCUS I44719 18 bp DNA
DEFINITION Sequence 45 from patent US 5635384.
ACCESSION I44719
VERSION I44719.1 GI:2469432
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 18)
AUTHORS Walsh,T.A., Hey,T.D. and Morgan,A.E.R.
TITLE Ribosome-inactivating proteins, inactive precursor forms thereof, a
process for making and a method of using
JOURNAL Patent: US 5635384-A 45 03-JUN-1997;
FEATURES
source
BASE COUNT 5 a 4 c 4 g 5 t
ORIGIN

Query Match 89.3%; Score 13.4; DB 10; Length 18;
Best Local Similarity 93.3%; Pred. No. 1.le+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 taagcagctgccatt 15
|||||
Db 17 TAAGCAGCTGCAATT 3

RESULT 2
152125/c
LOCUS I52125 18 bp DNA
DEFINITION Sequence 45 from patent US 5646026.
ACCESSION I52125
VERSION I52125.1 GI:2473326
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 18)
AUTHORS Walsh,T.A., Hey,T.D. and Morgan,A.E.R.
TITLE Ribosome-inactivating proteins, inactive precursor forms thereof, a
process for making and a method of using
JOURNAL Patent: US 5646026-A 45 08-JUL-1997;
FEATURES
source
BASE COUNT 5 a 4 c 4 g 5 t
ORIGIN

Query Match 89.3%; Score 13.4; DB 10; Length 18;
Best Local Similarity 93.3%; Pred. No. 1.le+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 taagcagctgccatt 15
|||||
Db 17 TAAGCAGCTGCAATT 3

RESULT 3
AR032874
LOCUS AR032874 50 bp DNA
DEFINITION Sequence 486 from patent US 5869241.
ACCESSION AR032874
VERSION AR032874.1 GI:5948479
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 50)
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.
TITLE Method of determining DNA sequence preference of a DNA-binding
molecule
JOURNAL Patent: US 5869241-A 486 09-FEB-1999;
FEATURES
source
BASE COUNT 9 a 14 c 13 g 14 t
ORIGIN

Query Match 78.7%; Score 11.8; DB 9; Length 50;
Best Local Similarity 86.7%; Pred. No. 1.4e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 taagcagctgccatt 15
|||||
Db 24 TAAGCAGCTGCTTTT 38

RESULT 4
AR032875
LOCUS AR032875 50 bp DNA
DEFINITION Sequence 487 from patent US 5869241.
ACCESSION AR032875
VERSION AR032875.1 GI:5948480
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 50)

```

AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule

JOURNAL Patent: US 5869241-A 487 09-FEB-1999;

FEATURES Location/Qualifiers  
1..50

BASE COUNT 8 a 13 c 13 g 16 t

ORIGIN

Query Match 78.7%; Score 11.8; DB 9; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.4e+04;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 5

I29614

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 6

I29615

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule

JOURNAL Patent: US 5869241-A 487 09-FEB-1999;

FEATURES Location/Qualifiers  
1..50

BASE COUNT 8 a 13 c 13 g 16 t

ORIGIN

Query Match 78.7%; Score 11.8; DB 9; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.4e+04;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 5

I29614

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 6

I29615

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule

JOURNAL Patent: US 5869241-A 487 09-FEB-1999;

FEATURES Location/Qualifiers  
1..50

BASE COUNT 8 a 13 c 13 g 16 t

ORIGIN

Query Match 78.7%; Score 11.8; DB 9; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.4e+04;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 5

I29614

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 6

I29615

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule

JOURNAL Patent: US 5869241-A 487 09-FEB-1999;

FEATURES Location/Qualifiers  
1..50

BASE COUNT 8 a 13 c 13 g 16 t

ORIGIN

Query Match 78.7%; Score 11.8; DB 9; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.4e+04;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 5

I29614

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 6

I29615

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule

JOURNAL Patent: US 5869241-A 487 09-FEB-1999;

FEATURES Location/Qualifiers  
1..50

BASE COUNT 8 a 13 c 13 g 16 t

ORIGIN

Query Match 78.7%; Score 11.8; DB 9; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.4e+04;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 5

I29614

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 6

I29615

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..50

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 13; Conservative

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule

JOURNAL Patent: US 5869241-A 487 09-FEB-1999;

FEATURES Location/Qualifiers  
1..50

BASE COUNT 8 a 13 c 13 g 16 t

ORIGIN

Query Match 78.7%; Score 11.8; DB 9; Length 50;

Best Local Similarity 86.7%; Pred. No. 1.4e+04;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taagcagctgccatt 15

|||||

Db 24 TAAGCAGCTGCTTTT 38

RESULT 5

I29614

LOCUS

DEFINITION

**KEYWORDS**  
SOURCE  
ORGANISM

**REFERENCE**  
**AUTHORS**

**TITLE**

**JOURNAL**

**FEATURES**  
source

**BASE COUNT**  
ORIGIN

Query Match 76.0%; Score 11.4; DB 10; Length 21;  
Best Local Similarity 80.0%; Pred. No. 2.5e+04;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagtgcatt 15  
|||||  
Db 17 TAGGCARCTGGATT 3

**RESULT 10**  
AB0791/c  
LOCUS AR0791 22 bp DNA PAT 21-JAN-2000  
DEFINITION Sequence 46 from Patent EP0922767.  
ACCESSION AR0791  
VERSION AR0791.1 GI:6731415  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.

**REFERENCE**  
1 (bases 1 to 22)  
Tucker, W. and Huttner, E.  
Ribbonzymes capable of conferring resistance to potyvirus infection,  
and plants expressing said ribozymes  
Patent: EP 0922767-A 46 16-JUN-1999;  
GENE SHEARS PTY LTD (AU)

**FEATURES**  
source Location/Qualifiers  
1..22 /organism="unidentified"  
/db\_xref="taxon:32644"

**BASE COUNT**  
ORIGIN

Query Match 76.0%; Score 11.4; DB 9; Length 22;  
Best Local Similarity 92.3%; Pred. No. 2.5e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 agcagctgacctt 15  
|||||  
Db 13 AGCAGCTGCAATT 1

**RESULT 11**  
AX022755/c  
LOCUS AX022755 22 bp DNA PAT 07-SEP-2000  
DEFINITION Sequence 46 from Patent WO9928485.  
ACCESSION AX022755  
VERSION AX022755.1 GI:10046280  
KEYWORDS unclassified.  
SOURCE unclassified.

**REFERENCE**  
1 (bases 1 to 22)  
Tucker, W. and Huttner, E.  
Ribbonzymes capable of conferring resistance to potyvirus infection,  
and plants expressing said ribozymes  
Patent: WO 9928485-A 46 10-JUN-1999;  
GENE SHEARS PTY LTD (AU); HUTTNER ERIC (FR);  
IGNART FREDERIC (FR); VERMEULEN AGNES (FR)

**FEATURES**  
source Location/Qualifiers  
1..22 /organism="unidentified"  
/db\_xref="taxon:32644"

**BASE COUNT**  
ORIGIN

Query Match 76.0%; Score 11.4; DB 9; Length 22;  
Best Local Similarity 92.3%; Pred. No. 2.5e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 agcagctgacctt 15  
|||||  
Db 13 AGCAGCTGCAATT 1

**RESULT 12**  
AR016058  
LOCUS AR016058 25 bp DNA PAT 05-DEC-1998  
DEFINITION Sequence 26 from patent US 5776679.  
ACCESSION AR016058  
VERSION AR016058.1 GI:3972335  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

**REFERENCE**  
1 (bases 1 to 25)  
Vilpeponteau, B., Feng, J., Funk, W. and Andrews, W.H.  
Assays for the DNA component of human telomerase  
Patent: US 5776679-A 26 07-JUL-1998;

**FEATURES**  
source Location/Qualifiers  
1..25 /organism="unknown"

**BASE COUNT**  
ORIGIN

Query Match 76.0%; Score 11.4; DB 9; Length 25;  
Best Local Similarity 92.3%; Pred. No. 2.5e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 agcagctgacctt 15  
|||||  
Db 13 AGCAGCTGCAATT 25

**RESULT 13**  
AR075530  
LOCUS AR075530 25 bp DNA PAT 30-AUG-2000  
DEFINITION Sequence 27 from patent US 5958680.  
ACCESSION AR075530  
VERSION AR075530.1 GI:10002278  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

**REFERENCE**  
1 (bases 1 to 25)  
Vilpeponteau, B., Feng, J., Funk, W. and Andrews, W.H.  
Mammalian telomerase  
Patent: US 5958680-A 27 28-SEP-1999;

**FEATURES**  
source Location/Qualifiers  
1..25 /organism="unknown"

**BASE COUNT**  
ORIGIN

AUTHORS	Tucker,W., Huttner,E., Ignart,F. and Vermeulen,A.
TITLE	Ribozymes capable of conferring resistance to potyvirus infection, and plants expressing said ribozymes
JOURNAL	Patent: WO 9928485-A 46 10-JUN-1999; GENE SHEARS PTY LTD (AU); TUCKER WILLIAM (AU); HUTTNER ERIC (FR); IGNART FREDERIC (FR); VERMEULEN AGNES (FR)
FEATURES	Location/Qualifiers source 1..22 /organism="unidentified" misc_feature 1..22 /db_xref="taxon:32644"
BASE COUNT	5 a 5 c 6 g 6 t
ORIGIN	
Query Match	76.0%; Score 11.4; DB 9; Length 22;
Best Local Similarity	92.3%; Pred. No. 2.5e+04;
Matches	12; Conservative 0; Mismatches 1; Indels 0; Gaps 0
Qy	3 agcagctgccatt 15 
Db	13 AGCAGCTGCAATT 1
RESULT 12	
LOCUS	AR016058 25 bp DNA PAT 05-DEC-1998
DEFINITION	Sequence 26 from patent US 5776679.
ACCESSION	AR016058
VERSION	AR016058.1 GI:3972335
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 25)
AUTHORS	Villepoiteau,B., Feng,J., Funk,W. and Andrews,W.H.
TITLE	Assays for the DNA component of human telomerase
JOURNAL	Patent: US 5776679-A 26 07-JUL-1998;
FEATURES	Location/Qualifiers source 1..25 /organism="unknown"
BASE COUNT	8 a 8 c 6 g 3 t
ORIGIN	
Query Match	76.0%; Score 11.4; DB 9; Length 25;
Best Local Similarity	92.3%; Pred. No. 2.5e+04;
Matches	12; Conservative 0; Mismatches 1; Indels 0; Gaps 0
Qy	3 agcagctgccatt 15 
Db	13 AGCAGCTGCAATT 25
RESULT 13	
LOCUS	AR075530 25 bp DNA PAT 30-AUG-2000
DEFINITION	Sequence 27 from patent US 5958680.
ACCESSION	AR075530
VERSION	AR075530.1 GI:10002278
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 25)
AUTHORS	Villepoiteau,B., Feng,J., Funk,W. and Andrews,W.H.
TITLE	Mammalian telomerase
JOURNAL	Patent: US 5958680-A 27 28-SEP-1999;
FEATURES	Location/Qualifiers source 1..25 /organism="unknown"
BASE COUNT	8 a 8 c 6 g 3 t

KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 21)
AUTHORS	Lander,E.S., Gargill,M., Ireland,J.S., Bolk,S., Daley,G.O. and McCarthy,J.J.
TITLE	Single nucleotide polymorphisms in genes
JOURNAL	Patent: WO 0118250-A 23 15-MAR-2001; WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH (US) ; Millennium Pharmaceuticals, Inc. (US)
FEATURES	Location/Qualifiers source 1..21 /organism="Homo sapiens" BASE COUNT 6 a 6 c 4 g 4 t 1 others
ORIGIN	
Query Match	76.0%; Score 11.4; DB 10; Length 21;
Best Local Similarity	80.0%; Pred. No. 2.5e+04;
Matches	12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy	1 taagcagtgcatt 15 
Db	17 TAGGCARCTGGATT 3
RESULT 10	
LOCUS	AB0791/c 22 bp DNA PAT 21-JAN-2000
DEFINITION	Sequence 46 from Patent EP0922767.
ACCESSION	AB0791
VERSION	AB0791.1 GI:6731415
KEYWORDS	unidentified.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 22)
AUTHORS	Tucker,W. and Huttner,E.
TITLE	Ribozymes capable of conferring resistance to potyvirus infection, and plants expressing said ribozymes
JOURNAL	Patent: EP 0922767-A 46 16-JUN-1999; GENE SHEARS PTY LTD (AU)
FEATURES	Location/Qualifiers source 1..22 /organism="unidentified" BASE COUNT 5 a 5 c 6 g 6 t
ORIGIN	
Query Match	76.0%; Score 11.4; DB 9; Length 22;
Best Local Similarity	92.3%; Pred. No. 2.5e+04;
Matches	12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	3 agcagctgccatt 15 
Db	13 AGCAGCTGCAATT 1
RESULT 11	
LOCUS	AX022755/c 22 bp DNA PAT 07-SEP-2000
DEFINITION	Sequence 46 from Patent WO9928485.
ACCESSION	AX022755
VERSION	AX022755.1 GI:10046280
KEYWORDS	unidentified.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 22)

## ORIGIN

Query Match 76.0%; Score 11.4; DB 9; Length 25;  
Best Local Similarity 92.3%; Pred. No. 2.5e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 agcagctgccatt 15  
|||||  
Db 13 AGCAGCTGACATT 25

## RESULT 14

AX019566  
LOCUS AX019566 25 bp DNA PAT 07-SEP-2000  
DEFINITION Sequence 20 from Patent WO9938964.  
ACCESSION AX019566

VERSION AX019566.1 GI:10043480

## KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 25)

AUTHORS Keith,W.N.

TITLE Promoter regions of the mouse and human telomerase rna component

genes

JOURNAL Patent: WO 9938964-A 20 05-AUG-1999;

KEITH WILLIAM NICOL (GB); CANCER RES CAMPAIGN TECH (GB)

## FEATURES

source Location/Qualifiers

1..25

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer"

6 a 7 c 9 g 3 t

## BASE COUNT

ORIGIN

Query Match 76.0%; Score 11.4; DB 9; Length 25;  
Best Local Similarity 92.3%; Pred. No. 2.5e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 agcagctgccatt 15  
|||||  
Db 13 AGCAGCTGACATT 25

## RESULT 15

E04774/c

## LOCUS

DEFINITION E04774 26 bp DNA PAT 29-SEP-1997  
Synthetic DNA for site directed mutagenesis of protease derived

from Bacillus NKS-21.

ACCESSION E04774

VERSION E04774.1 GI:2172970

KEYWORDS JP 1993091876-A/13.

SOURCE synthetic construct

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 26)

AUTHORS Takinishi,E., Kakinuma,S., Takemoto,A., Miyoda,Y. and Fukuyama,S.

TITLE PROTEASE, ITS PRODUCTION AND USE

JOURNAL Patent: JP 1993091876-A 13 16-APR-1993;

SHOWA DENKO KK

## COMMENT

OS Artificial gene

OC Artificial sequence; Genes.

OS Bacillus NKS-21

PN JP 1993091876-A/13

PD 16-APR-1993

PF 02-OCT-1991 JP 1991280313

PI TAKINISHI EIKO, KAKINUMA SHINJI, TAKEMOTO AYANO, PI MIYODA

YOSHIAKI,

PI FUKUYAMA SHIRO

PC C12N9/54,C11D3/386,C12N1/21,C12N15/57,(C12N9/54,C12R1:07), PC

(C12N1/21,

PC C12R1:07);  
CC strandedness: Single;  
CC topology: Linear;  
CC hypothetical: NO;  
CC anti-sense: NO;  
FEATURES  
source Location/Qualifiers  
1..26  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
7 a 5 c 5 g 9 t

Query Match 76.0%; Score 11.4; DB 10; Length 26;  
Best Local Similarity 92.3%; Pred. No. 2.5e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 agcagctgccatt 15  
|||||  
Db 18 AGCAGCTGCAATT 6

Search completed: October 2, 2001, 15:56:53  
Job time: 14177 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:11 ; Search time 10798.2 Seconds  
(without alignments)  
13.131 Million cell updates/sec

Title: US-09-757-100B-30

Perfect score: 15

Sequence: 1 taagcagctgccatt 15

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*\*

- 1: gb\_est1:\*\*
- 2: gb\_est2:\*\*
- 3: gb\_est3:\*\*
- 4: gb\_est4:\*\*
- 5: gb\_est5:\*\*
- 6: gb\_est6:\*\*
- 7: gb\_est7:\*\*
- 8: gb\_est8:\*\*
- 9: gb\_est9:\*\*
- 10: gb\_est10:\*\*
- 11: gb\_est11:\*\*
- 12: gb\_est12:\*\*
- 13: gb\_est13:\*\*
- 14: gb\_est14:\*\*
- 15: gb\_est15:\*\*
- 16: gb\_est16:\*\*
- 17: gb\_est17:\*\*
- 18: gb\_est18:\*\*
- 19: gb\_est19:\*\*
- 20: gb\_est20:\*\*
- 21: gb\_est21:\*\*
- 22: gb\_est22:\*\*
- 23: gb\_est23:\*\*
- 24: gb\_est24:\*\*
- 25: gb\_est33:\*\*
- 26: gb\_est34:\*\*
- 27: gb\_est35:\*\*
- 28: gb\_est36:\*\*
- 29: gb\_est37:\*\*
- 30: gb\_est38:\*\*
- 31: gb\_est39:\*\*
- 32: gb\_est40:\*\*
- 33: em\_estba:\*\*
- 34: em\_estfun:\*\*
- 35: em\_esthum1:\*\*
- 36: em\_esthum2:\*\*
- 37: em\_esthum3:\*\*
- 38: em\_esthum4:\*\*
- 39: em\_esthum5:\*\*
- 40: em\_esthum6:\*\*
- 41: em\_esthum7:\*\*
- 42: em\_esthum8:\*\*
- 43: em\_esthum9:\*\*

- 44: em\_esthum10:\*\*
- 45: em\_esthum11:\*\*
- 46: em\_esthum12:\*\*
- 47: em\_esthum13:\*\*
- 48: em\_esthum14:\*\*
- 49: em\_esthum15:\*\*
- 50: em\_esthum16:\*\*
- 51: em\_esthum17:\*\*
- 52: em\_esthum18:\*\*
- 53: em\_esthum19:\*\*
- 54: em\_esthum20:\*\*
- 55: em\_esthum21:\*\*
- 56: em\_esthum22:\*\*
- 57: em\_esthum23:\*\*
- 58: em\_esthum24:\*\*
- 59: em\_esthum25:\*\*
- 60: em\_esthum26:\*\*
- 61: em\_esthum27:\*\*
- 62: em\_esthum28:\*\*
- 63: em\_estin1:\*\*
- 64: em\_estin2:\*\*
- 65: em\_estin3:\*\*
- 66: em\_estin4:\*\*
- 67: em\_estin5:\*\*
- 68: em\_estom1:\*\*
- 69: em\_estom2:\*\*
- 70: em\_estov1:\*\*
- 71: em\_estov2:\*\*
- 72: em\_estpl1:\*\*
- 73: em\_estpl2:\*\*
- 74: em\_estpl3:\*\*
- 75: em\_estpl4:\*\*
- 76: em\_estpl5:\*\*
- 77: em\_estpl6:\*\*
- 78: em\_estpl7:\*\*
- 79: em\_estpl8:\*\*
- 80: em\_estpl9:\*\*
- 81: em\_estpl10:\*\*
- 82: em\_estro1:\*\*
- 83: em\_estro2:\*\*
- 84: em\_estro3:\*\*
- 85: em\_estro4:\*\*
- 86: em\_estro5:\*\*
- 87: em\_estro6:\*\*
- 88: em\_estro7:\*\*
- 89: em\_estro8:\*\*
- 90: em\_estro9:\*\*
- 91: em\_estro10:\*\*
- 92: em\_estro11:\*\*
- 93: em\_estro12:\*\*
- 94: em\_estro13:\*\*
- 95: em\_estro14:\*\*
- 96: em\_estro15:\*\*
- 97: em\_estro16:\*\*
- 98: em\_estro17:\*\*
- 99: em\_estro18:\*\*
- 100: em\_estro19:\*\*
- 101: em\_estro20:\*\*
- 102: gb\_est25:\*\*
- 103: gb\_est26:\*\*
- 104: gb\_est27:\*\*
- 105: gb\_est28:\*\*
- 106: gb\_est29:\*\*
- 107: gb\_est30:\*\*
- 108: gb\_est31:\*\*
- 109: gb\_est32:\*\*
- 110: gb\_est41:\*\*
- 111: gb\_est42:\*\*
- 112: gb\_est43:\*\*
- 113: gb\_est44:\*\*
- 114: gb\_est45:\*\*
- 115: gb\_est46:\*\*
- 116: gb\_est47:\*\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	11.8	78.7	41	12	AA829551	AA829551 of09h01.s
c 2	11.8	78.7	50	10	AA662608	AA662608 nr16a12.s
c 3	11.4	76.0	47	143	AA794942	AA794942 vr05g03.r
c 4	11.4	76.0	45	12	BF036368	BF036368 601460534
c 5	11	73.3	47	158	H59245	H59245 yr04b03.s1
c 6	11	73.3	50	107	AU103473	AU103473 AU103473
c 7	11	73.3	50	107	AU104774	AU104774 AU104774
c 8	10.8	72.0	19	244	AZ481008	AZ481008 IM0302N15
c 9	10.8	72.0	28	258	TA355F03Q	TA355F03Q T. brucei
c 10	10.8	72.0	40	12	AA806226	AA806226 oe29f12.s
c 11	10.8	72.0	41	258	TA126G07Q	TA126G07Q T. brucei
c 12	10.8	72.0	42	113	AW249855	AW249855 2821534.3
c 13	10.8	72.0	44	9	AA553766	AA553766 nk29h01.s
c 14	10.8	72.0	44	9	AA600013	AA600013 ag29h11.s
c 15	10.8	72.0	44	11	AA713885	AA713885 nv79d06.s
c 16	10.8	72.0	44	14	AA989114	AA989114 or75g04.s
c 17	10.8	72.0	50	20	AI476044	AI476044 t197b11.x
c 18	10.8	72.0	50	107	AU102666	AU102666 AU102666
c 19	10.8	72.0	50	107	AU103462	AU103462 AU103462
c 20	10.8	72.0	50	107	AU104066	AU104066 AU104066
c 21	10.8	72.0	50	107	AU107071	AU107071 AU107071
c 22	10.4	69.3	20	243	AZ435787	AZ435787 IM0223G02
c 23	10.4	69.3	21	242	AZ374998	AZ374998 IM0128H11
c 24	10.4	69.3	22	243	AZ430117	AZ430117 IM0214M24
c 25	10.4	69.3	22	249	AZ805989	AZ805989 2M0067C13
c 26	10.4	69.3	26	247	AZ632821	AZ632821 IM0487K12
c 27	10.4	69.3	31	15	AI032592	AI032592 ow73g08.s
c 28	10.4	69.3	31	244	AZ448456	AZ448456 IM0246H08
c 29	10.4	69.3	32	258	TA371C10P	TA371C10P T. brucei
c 30	10.4	69.3	34	4	AA282064	AA282064 zt02e01.s
c 31	10.4	69.3	34	249	AZ785014	AZ785014 2M0028D03
c 32	10.4	69.3	38	249	AZ759713	AZ759713 IM0552E14
c 33	10.4	69.3	41	241	AZ327077	AZ327077 IM0050C16
c 34	10.4	69.3	41	247	AZ634745	AZ634745 IM0490O12
c 35	10.4	69.3	42	249	AZ760286	AZ760286 IM0553M20
c 36	10.4	69.3	45	188	T17561	T17561 mps v160.th
c 37	10.4	69.3	48	249	AZ787799	AZ787799 2M0034C06
c 38	10.4	69.3	49	1	AA047951	AA047951 mj23h04.r
c 39	10.4	69.3	49	13	AA914273	AA914273 vY99a03.r
c 40	10.4	69.3	50	107	AU102448	AU102448 AU102448
c 41	10.4	69.3	50	107	AU103167	AU103167 AU103167
c 42	10.4	69.3	50	107	AU103390	AU103390 AU103390
c 43	10.4	69.3	50	107	AU103485	AU103485 AU103485
c 44	10.4	69.3	50	107	AU103486	AU103486 AU103486
c 45	10.4	69.3	50	107	AU103487	AU103487 AU103487

ALIGNMENTS

RESULT 1  
AA829551/c  
LOCUS  
DEFINITION  
OF09h01.s1 NCI\_CGAP\_Col2 Homo sapiens cdna clone IMAGE:1420657 3'  
similar to gb:M33552 LYMPHOCYTE-SPECIFIC PROTEIN LSP1 (HUMAN);,  
mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AA829551 41 bp mRNA EST 29-APR-1998  
OF09h01.s1 NCI\_CGAP\_Col2 Homo sapiens cdna clone IMAGE:1420657 3'  
similar to gb:M33552 LYMPHOCYTE-SPECIFIC PROTEIN LSP1 (HUMAN);,  
mRNA sequence.  
AA829551.1 GI:2902650  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 41)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Stratagene, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone Sequencing by: Washington University Genome Sequencing Center  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
Insert Length: 660 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES

source  
1..41  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1420657"  
/clone\_lib="NCI\_CGAP\_Col2"  
/sex="mixed"  
/tissue\_type="colon tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: colon; Vector: Bluescript SK-; Site:1: EcoRI  
; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT. Pooled colon tumors. 5' adaptor sequence: 5'  
GAATTCGCCACGAG 3' 3' adaptor sequence: 5'  
CTCGAGTGTGTTTTTTTTTTT 3' Average insert size: 1.2 kb."  
BASE COUNT 10 a 6 c 11 g 14 t  
ORIGIN

Query Match 78.7%; Score 11.8; DB 12; Length 41;  
Best Local Similarity 86.7%; Pred. No. 3.2e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 taagcagctgccatt 15  
||||||| |||||  
Db 19 TAAGCAGCCACCAT 5

RESULT 2

AA662608 50 bp mRNA EST 12-NOV-1997  
LOCUS  
DEFINITION  
nr16a12.s1 NCI\_CGAP\_Ew1 Homo sapiens cdna clone IMAGE:1168126  
similar to SW:RL24\_HUMAN P38663 60S RIBOSOMAL PROTEIN L24. ;, mRNA  
sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 50)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40ml3 fwd. ET from Amersham.  
Location/Qualifiers

```

source
1. .50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1168126"
/clone_lib="NCI_CGAP_Ew1"
/tissue_type="Ewing's sarcoma"
/lab_host="DH10B"
/Note="Vector: pAMP10; mRNA made from Ewing's sarcoma,
cDNA made by oligo-dT priming. Non-directionally cloned.
Size selected on agarose gel, average insert size 600 bp.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT      9 a      11 c      14 g      16 t
ORIGIN

Query Match      78.7%; Score 11.8; DB 10; Length 50;
Best Local Similarity 86.7%; Pred. No. 3.3e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15
    | | | | | | | | | |
Db 11 TTAGCAGCGCCATT 25

RESULT 3
AA794942/c
LOCUS      AA794942      45 bp      mRNA      EST      09-FEB-1998
DEFINITION      vr05g03.r1 Knowles Solter blastocyst B3 Mus musculus cDNA
                clone IMAGE:110964 5' similar to gb:M6124 Mus musculus EGP314
                precursor mRNA, complete cds (MOUSE);, mRNA sequence.
ACCESSION      AA794942
VERSION
KEYWORDS
SOURCE
ORGANISM      house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 45)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Thelshing,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:609132
Trace considered overall poor quality
High quality sequence stop: 1.
FEATURES
source
1. 45
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:110964"
/clone_lib="Knowles Solter mouse blastocyst B3"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/Note="Organ: embryo; Vector: pSPORT; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dT):
5'-CGGTGACCGTCGACCGTTTTTTTTTT-3', cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life

```

```

Technologies). Two different size selections: B1 (larger
inserts) and B3,
BASE COUNT      14 a      8 c      10 g      13 t
ORIGIN

Query Match      76.0%; Score 11.4; DB 12; Length 45;
Best Local Similarity 92.3%; Pred. No. 5.4e+04;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 taagcagctgccca 13
    | | | | | | | | | |
Db 44 TTAGCAGCTGCCA 32

RESULT 4
BF036368/c
LOCUS      BF036368      47 bp      mRNA      EST      20-OCT-2000
DEFINITION      601460534F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3863768 5',
                mRNA sequence.
ACCESSION      BF036368
VERSION
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 47)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9604 row: d column: 09
High quality sequence stop: 47.
FEATURES
source
1. 47
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3863768"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: ovary; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT      5 a      20 c      15 g      7 t
ORIGIN

Query Match      76.0%; Score 11.4; DB 143; Length 47;
Best Local Similarity 92.3%; Pred. No. 5.4e+04;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 aagcagctgccat 14
    | | | | | | | | | |
Db 41 AAGCGGCTGCCAT 29

RESULT 5
H59245
LOCUS      H59245      47 bp      mRNA      EST      06-OCT-1995
DEFINITION      yr04b03.s1 Soares fetal liver spleen cDNA clone
                IMAGE:204269 3' similar to gb:M61906 PHOSPHATIDYLINOSITOL 3-KINASE
                REGULATORY ALPHA SUBUNIT (HUMAN);, mRNA sequence.

```

```

ACCESSION H59245
VERSION H59245.1 GI:1012077
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman
M., Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,
Rifkin L., Rohlfing T., Soares M., Tan F., Trevaskis E., Waterston
R., Williamson A., Wohlmann P. and Wilson R.
The WashU-Merck EST Project
Contact: Wilson RK
Unpublished (1995)
TITLE Washington University School of Medicine
JOURNAL 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1419
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 1419 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 1.

FEATURES
source
1. .50
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP17451"
/clone_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 8 a 12 c 11 g 19 t
ORIGIN

Query Match 73.3%; Score 11; DB 107; Length 50;
Best Local Similarity 100.0%; Pred. No. 9e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagcagctgc 11
|||||
Db 24 TAAGCAGCTGC 34

RESULT 7
AUI04774/c
LOCUS AUI04774 50 bp mRNA EST 05-APR-2001
DEFINITION AUI04774 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HRC05287, mRNA sequence.
ACCESSION AUI04774
VERSION AUI04774.1 GI:13554295
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Suzuki Y., Tsunoda T., Tanaka T., Nakamura Y., Morishita S., Okubo
H., Ota T., Isogai T., Tanaka T., Nakamura Y., Morishita S., Okubo
K., Suyama A. and Sugano S.
Fine Structural analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
Unpublished (2001)
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yszuki@ims.u-tokyo.ac.jp
Suzuki Y., Yoshitomo-Nakagawa K., Maruyama K., Suyama A. and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

TITLE Location/Qualifiers
JOURNAL 1. .50
COMMENT /organism="Homo sapiens"
/clone="HRC05287"
/clone_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 5 a 21 c 15 g 9 t
ORIGIN

Query Match 73.3%; Score 11; DB 107; Length 50;

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman
M., Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,
Rifkin L., Rohlfing T., Soares M., Tan F., Trevaskis E., Waterston
R., Williamson A., Wohlmann P. and Wilson R.
The WashU-Merck EST Project
Contact: Wilson RK
Unpublished (1995)
TITLE Washington University School of Medicine
JOURNAL 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1419
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 1419 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 1.

FEATURES
source
1. .47
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:3773400"
/db_xref="taxon:9606"
/clone="IMAGE:204269"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGTGAGAAATTAATTAAGATCTTTTGTGTTTGTGTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaído."
BASE COUNT 12 a 12 c 7 g 12 t 4 others
ORIGIN

Query Match 73.3%; Score 11; DB 159; Length 47;
Best Local Similarity 91.7%; Pred. No. 8.9e+04;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aagcagctgcc 13
|||||
Db 15 AAGCAGCTNCCA 26

RESULT 6
AUI03473
LOCUS AUI03473 50 bp mRNA EST 05-APR-2001
DEFINITION AUI03473 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HEP17451, mRNA sequence.
ACCESSION AUI03473
VERSION AUI03473.1 GI:13552994
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Best Local Similarity 100.0%; Pred. No. 9e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 aagcagctgcc 12
    ||| ||| ||| |||
Db 36 AAGCAGCTGCC 26

RESULT 8
LOCUS A2481008 19 bp DNA GSS 04-OCT-2000
DEFINITION IM0302N15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0302N15 R, DNA sequence.
ACCESSION A2481008
VERSION A2481008.1 GI:10641989
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 19)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relilly
, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 309, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0302 row: N column: 15
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 19.
FEATURES
Location/Qualifiers
1..19
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0302N15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 3 a 4 c 6 g 6 t
ORIGIN

Query Match 72.0%; Score 10.8; DB 244; Length 19;
Best Local Similarity 85.7%; Pred. No. 9.8e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aagcagctgccatt 15
    ||| ||| ||| ||| |||
Db 14 AAGCAGCTACCGTT 1

RESULT 9
LOCUS TA355F03Q/c 28 bp DNA GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 355f03, reverse sequence,
genomic survey sequence.
ACCESSION AL497072
VERSION AL497072.1 GI:11872664
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 28)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
Location/Qualifiers
1..28
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="355f03"
BASE COUNT 13 a 5 c 4 g 6 t
ORIGIN

Query Match 72.0%; Score 10.8; DB 258; Length 28;
Best Local Similarity 85.7%; Pred. No. 1e+05;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 14
    || ||| ||| ||| |||
Db 17 TATGCAGTTGCCAT 4

RESULT 10
LOCUS AA806226/c 40 bp mRNA EST 12-FEB-1998
DEFINITION Oe29f12.sl NCLCGAP_Pr25 Homo sapiens cDNA clone IMAGE:1409999 3'
similar to gb:M33552 LYMPHOCYTE-SPECIFIC PROTEIN LSPI (HUMAN);,
mRNA sequence.
ACCESSION AA806226
VERSION AA806226.1 GI:2874976
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE 1 (bases 1 to 40)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright, Ph.D.  
 CDNA Library Preparation: Stratagene, Inc.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 1.

## FEATURES

source  
 1..40  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1409999"  
 /clone\_lib="NCI\_CGAP\_Pr25"  
 /tissue\_type="epithelium (cell line)"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: prostate; Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Normal prostate epithelial cell line (HPV immortalized). 5' adaptor sequence: 5' GAATTCGCACGAC 3' 3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3' Average insert size: 1.1 kb."  
 7 a 6 c 11 g 16 t

## BASE COUNT

Query Match 72.0%; Score 10.8; DB 12; Length 40;  
 Best Local Similarity 85.7%; Pred. No. 1.1e+05;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aagcagctgccatt 15  
 ||||| |||||  
 Db 21 AAGCAGCCACCATT 8

## RESULT 11

TA126G07Q 41 bp DNA GSS 13-DEC-2000  
 LOCUS T. brucei sheared genomic DNA clone 126g07, reverse sequence, genomic survey sequence.  
 DEFINITION  
 AL463426  
 ACCESSION  
 VERSION AL463426.1 GI:11833936  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

## REFERENCE

1 (bases 1 to 41)  
 Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajadream, M.A. and Barrell, B.G.  
 Direct Submission  
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: [barrell@sanger.ac.uk](mailto:barrell@sanger.ac.uk) and [nhl@sanger.ac.uk](mailto:nhl@sanger.ac.uk)

## TITLE

Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

Email: [neilsayed@tigr.org](mailto:neilsayed@tigr.org)  
 Details of T. brucei sequencing at the Sanger Centre are available at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

## FEATURES

source  
 1..41  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="126g07"  
 14 a 9 c 6 g 12 t

## BASE COUNT

Query Match 72.0%; Score 10.8; DB 258; Length 41;  
 Best Local Similarity 85.7%; Pred. No. 1.1e+05;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taagcagctgccatt 14  
 ||||| |||||  
 Db 6 TAAGTAGATGCCAT 19

## RESULT 12

AW249855/c 42 bp mRNA EST 07-JAN-2000  
 LOCUS 2821534.3prime NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:2821534 3', mRNA sequence.  
 DEFINITION  
 AW249855  
 ACCESSION  
 VERSION AW249855.1 GI:6592848  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 42)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Other ESTs: 2821534.5prime  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: DCTD/DTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing project Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html) Base Calling / Quality Scores: PHRED from University of Washington Genome Center. Vector Trimming: cross\_match from University of Washington Genome Center. PHRAP suite. Poly-T Identification: patmatch.pl from Berkeley Drosophila Genome Project. University of Washington Genome Center: <http://www.genome.washington.edu/LowQualitySequence>: 10 contiguous PHRED high quality bases followed by vector sequence. Very Low Quality Sequence: Trace file contained 42 contiguous distinct peaks following vector sequence. Polyadenylation: Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this cDNA insert was polyadenylated.

Plate: LLCM7 row: A column: 23  
 High quality sequence stop: 10.

## FEATURES

source  
 1..42  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2821534"  
 /clone\_lib="NIH\_MGC\_7"  
 /tissue\_type="small cell carcinoma"  
 /cell\_line="MGC3"  
 /lab\_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 6 a 7 c 11 g 18 t

ORIGIN

Query Match 72.0%; Score 10.8; DB 113; Length 42;  
Best Local Similarity 85.7%; Pred. No. 1.1e+05;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aagcagctgccatt 15  
||||| |||  
Db 26 AAGCAGCTCCCTT 13

RESULT 13  
AA553766/c  
LOCUS  
DEFINITION  
AA553766 44 bp mRNA EST 08-SEP-1997  
nk29h01.s1 NCI\_CGAP\_Coll1 Homo sapiens cDNA clone IMAGE:1014961 3'  
similar to gb:M33552 LYMPHOCYTE-SPECIFIC PROTEIN LSP1 (HUMAN);  
mRNA sequence.

ACCESSION  
AA553766  
VERSION  
AA553766.1 GI:2324305  
KEYWORDS  
EST.  
SOURCE  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 44)  
AUTHORS  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Elias Campo,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Stratagene, Inc.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 865 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.

FEATURES  
source  
1. .44  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1014961"  
/clone\_lib="NCI\_CGAP\_Coll1"  
/tissue\_type="tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: colon; Vector: Bluescript SK-; Site\_1: EcoRI  
; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dt. Multiple colon tumors. 5' adaptor sequence: 5'  
GAATTCGCGCAGAG 3' 3' adaptor sequence: 5'  
CTCAGATTTT TTTT TTTT 3' Average insert size: 1.1 kb."

BASE COUNT 9 a 7 c 11 g 17 t

ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 44;  
Best Local Similarity 85.7%; Pred. No. 1.1e+05;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aagcagctgccatt 15  
||||| |||  
Db 21 AAGCAGCTCCCAAT 8

RESULT 14  
AA600013/c  
LOCUS  
DEFINITION  
AA600013 44 bp mRNA EST 12-MAY-1999  
ag29h11.s1 Jia bone marrow stroma Homo sapiens cDNA clone  
IMAGE:1091013 3' similar to gb:M33552 LYMPHOCYTE-SPECIFIC PROTEIN  
LSP1 (HUMAN); mRNA sequence.

ACCESSION  
AA600013  
VERSION  
AA600013.1 GI:2433638  
KEYWORDS  
EST.  
SOURCE  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 44)  
AUTHORS  
Jia, L., Robey, P., Young, M., Schuler, G., Powell, J., Yang, L., Lennon  
G., Hillier, L., Allen, M., Bowles, L., Geisel, S., Kucaba, T., Marra  
M., Martin, J., Steptoe, M., Tan, F., Theising, B., Bowers, Y., Wylie  
T., Waterston, R., Wilson, R. and Francomano, C.

TITLE  
WashU-MGB/NHGRI EST Project  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Wilton RK / Jia L  
WashU-MGB/NHGRI EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.

FEATURES  
Location/Qualifiers

1. .44  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1091013"  
/clone\_lib="Jia bone marrow stroma"  
/sex="mixed"  
/tissue\_type="bone marrow stroma"  
/dev\_stage="mixed"  
/lab\_host="XLI-Blue MRF/SOLR"  
/note="Vector: pBluescript; Site\_1: EcoRI; Site\_2: XhoI;  
mRNA made from human bone marrow stroma, cDNA made by  
oligo-dt priming. Directionally cloned. Size-selected for  
average insert size >0.5 kb. Library supplied by Dr. Libin  
Jia (NHGRI)."

BASE COUNT 9 a 7 c 11 g 17 t

ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 44;  
Best Local Similarity 85.7%; Pred. No. 1.1e+05;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aagcagctgccatt 15  
||||| |||  
Db 21 AAGCAGCCACCATT 8

RESULT 15  
AA713885/c  
LOCUS  
DEFINITION  
AA713885 44 bp mRNA EST 18-FEB-1998  
nv79d06.s1 NCI\_CGAP\_Br4 Homo sapiens cDNA clone IMAGE:1236011  
similar to gb:M33552 LYMPHOCYTE-SPECIFIC PROTEIN LSP1 (HUMAN);  
mRNA sequence.  
AA713885  
ACCESSION  
AA713885.1 GI:2726159  
VERSION  
AA713885  
KEYWORDS  
EST.



SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 44)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/hcicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-femail.nih.gov](mailto:cgaps-femail.nih.gov)  
Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,  
Ph.D. student, Rodrigo F. Chuqui, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
Insert Length: 521 Std Error: 0.00  
Seq primer: -40ml3 fwd. Et from Amersham  
High quality sequence stop: 1.  
FEATURES  
source  
1..44  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1236011"  
/clone\_lib="NCI\_CGAP\_Br4"  
/sex="female"  
/tissue\_type="normal ductal tissue"  
/lab\_host="DH10B"  
/note="Organ: breast; Vector: pAMP10; mRNA made from  
normal breast ductal tissue, cDNA made by oligo-dT  
priming. Non-directionally cloned. Size-selected on  
agarose gel, average insert size 600 bp. Reference:  
Krizman et al. (1996) Cancer Research 56:5380-5383."  
BASE COUNT 9 a 6 c 11 g 18 t  
ORIGIN

Query Match 72.0%; Score 10.8; DB 11; Length 44;  
Best Local Similarity 85.7%; Pred. No. 1.1e+05;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 aagcagctgccatt 15  
||||| |||||  
Db 21 AAGCAGCCACCATT 8

Search completed: October 2, 2001, 15:01:12  
Job time: 10841 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:53 ; Search time 3339.34 Seconds  
(without alignments)  
69.480 Million cell updates/sec

Title: US-09-757-100B-31  
Perfect score: 15  
Sequence: 1 taccagtgagtct 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 423528

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_ba3:\*
- 4: gb\_in1:\*
- 5: gb\_in2:\*
- 6: gb\_in3:\*
- 7: gb\_om:\*
- 8: gb\_ov:\*
- 9: gb\_pat1:\*
- 10: gb\_pat2:\*
- 11: gb\_ph:\*
- 12: gb\_pl1:\*
- 13: gb\_pl2:\*
- 14: gb\_pl3:\*
- 15: gb\_pl4:\*
- 16: em\_ba1:\*
- 17: em\_ba2:\*
- 18: em\_fun:\*
- 19: em\_htgo\_hum:\*
- 20: em\_htgo\_inv:\*
- 21: em\_htgo\_rod:\*
- 22: em\_htg\_hum1:\*
- 23: em\_htg\_hum2:\*
- 24: em\_htg\_hum3:\*
- 25: em\_htg\_hum4:\*
- 26: em\_htg\_hum5:\*
- 27: em\_htg\_hum6:\*
- 28: em\_htg\_hum7:\*
- 29: em\_htg\_hum8:\*
- 30: em\_htg\_inv1:\*
- 31: em\_htg\_inv2:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_rod:\*
- 34: em\_hum1:\*
- 35: em\_hum2:\*
- 36: em\_hum3:\*
- 37: em\_hum4:\*
- 38: em\_hum5:\*
- 39: em\_hum6:\*
- 40: em\_hum7:\*
- 41: em\_in:\*
- 42: em\_om:\*
- 43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_rod:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vi:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vil:\*

59: gb\_v12:\*

60: gb\_htg1:\*

61: gb\_htg2:\*

62: gb\_htg3:\*

63: gb\_htg4:\*

64: gb\_htg5:\*

65: gb\_htg6:\*

66: gb\_htg7:\*

67: gb\_htg8:\*

68: gb\_htg9:\*

69: gb\_htg10:\*

70: gb\_htg11:\*

71: gb\_htg12:\*

72: gb\_htg13:\*

73: gb\_htg14:\*

74: gb\_htg15:\*

75: gb\_htg16:\*

76: gb\_htg17:\*

77: gb\_htg18:\*

78: gb\_htg19:\*

79: gb\_htg20:\*

80: gb\_htg21:\*

81: gb\_htg22:\*

82: gb\_htg23:\*

83: gb\_htg24:\*

84: gb\_htg25:\*

85: gb\_pr1:\*

86: gb\_pr2:\*

87: gb\_pr3:\*

88: gb\_pr4:\*

89: gb\_pr5:\*

90: gb\_pr6:\*

91: gb\_pr7:\*

92: gb\_pr8:\*

93: gb\_pr9:\*

94: gb\_rol:\*

95: gb\_ro2:\*

96: gb\_in4:\*

97: gb\_pr10:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	12.4	82.7	21	10	I30544
C 2	12.4	82.7	22	9	AR067320
C 3	11.8	78.7	34	9	AR091705
C 4	11.8	78.7	48	9	AR075823
C 5	11.8	78.7	48	10	E30456
C 6	11.4	76.0	18	9	A34802
C 7	11.4	76.0	30	10	I36150
C 8	11	73.3	20	9	AR060552
					I30544 Sequence 7
					AR067320 Sequence
					AR091705 Sequence
					AR075823 Sequence
					E30456 Method for
					A34802 HSV probe.
					I36150 Sequence 34
					AR060552 Sequence

```

9      11 73.3 20 10 AX069241 Sequence
10     11 73.3 28 9  AR003365 Sequence
11     11 73.3 28 9  AR067869 Sequence
12     10.8 72.0 21 9  AR043988 Sequence
13     10.8 72.0 21 9  AR043989 Sequence
14     10.8 72.0 21 9  AR073521 Sequence
15     10.8 72.0 21 9  AR073522 Sequence
16     10.8 72.0 21 9  AR022131 Sequence
17     10.8 72.0 21 9  AX022132 Sequence
18     10.8 72.0 21 10 193392 Sequence
19     10.8 72.0 21 10 193393 Sequence
20     10.8 72.0 26 9  AR013888 Sequence
21     10.8 72.0 26 9  AR033842 Sequence
22     10.8 72.0 26 9  AR042502 Sequence
23     10.8 72.0 26 9  AR058382 Sequence
24     10.8 72.0 26 9  AR089208 Sequence
25     10.8 72.0 26 10 175320 Sequence
26     10.8 72.0 36 94 M0SIGHID2
27     10.8 72.0 37 9  AX010786 Sequence
28     10.8 72.0 40 9  AR013900 Sequence
29     10.8 72.0 40 9  AR033854 Sequence
30     10.8 72.0 40 9  AR042514 Sequence
31     10.8 72.0 40 9  AR058394 Sequence
32     10.8 72.0 40 9  AR088220 Sequence
33     10.8 72.0 42 9  A36516 Sequence
34     10.8 72.0 42 9  AR013872 Sequence
35     10.8 72.0 42 9  AR033826 Sequence
36     10.8 72.0 42 9  AR042486 Sequence
37     10.8 72.0 42 9  AR058366 Sequence
38     10.8 72.0 42 9  AR080149 Sequence
39     10.8 72.0 42 9  AR088192 Sequence
40     10.8 72.0 44 9  AR035220 Sequence
41     10.8 72.0 44 9  AR035241 Sequence
42     10.8 72.0 50 10 143338 Sequence
43     10.8 72.0 50 10 143338 Sequence
44     10.4 69.3 15 9  AX018749 Sequence
45     10.4 69.3 16 9  AX015626 Sequence

ALIGNMENTS

RESULT 1
130544/c
LOCUS      I30544      21 bp      DNA
DEFINITION Sequence 7 from patent US 5580969.
ACCESSION  I30544
VERSION     I30544.1 GI:1821335
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 21)
AUTHORS    Hoke,G.D., Bradley,M.O., Williams,T.J. and Lee,C.
TITLE      Antisense oligonucleotides directed against human ICAM-1 RNA
JOURNAL    Patent: US 5580969-A 7 03-DEC-1996;
FEATURES   Location/Qualifiers
            1..21
            /organism="unknown"
BASE COUNT 5 a 9 c 5 g 2 t
ORIGIN

Query Match      82.7%; Score 12.4; DB 10; Length 21;
Best Local Similarity 92.9%; Pred. No. 5.4e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 taccagggtgagtc 14
Db 18 TTCCAGGTGAGTC 5

RESULT 2

AX069241/c
LOCUS      AR067320      22 bp      DNA
DEFINITION Sequence 668 from patent US 5851760.
ACCESSION  AR067320
VERSION     AR067320.1 GI:5998542
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 22)
AUTHORS    Evans,G.A. and Smith,M.W.
TITLE      Method for generation of sequence sampled maps of complex genomes
JOURNAL    Patent: US 5851760-A 668 22-DEC-1998;
FEATURES   Location/Qualifiers
            1..22
            /organism="unknown"
BASE COUNT 8 a 5 c 4 g 5 t
ORIGIN

Query Match      82.7%; Score 12.4; DB 9; Length 22;
Best Local Similarity 92.9%; Pred. No. 5.4e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 accagggtgagtc 15
Db 14 ATCCAGGTGAGTCT 1

RESULT 3
AR091705
LOCUS      AR091705      34 bp      DNA
DEFINITION Sequence 15 from patent US 5994505.
ACCESSION  AR091705
VERSION     AR091705.1 GI:10018459
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 34)
AUTHORS    Ting,J.Pan-Yung and Chin,K.
TITLE      Forms of class II MHC transactivator (CIITA)
JOURNAL    Patent: US 5994505-A 15 30-NOV-1999;
FEATURES   Location/Qualifiers
            1..34
            /organism="unknown"
BASE COUNT 7 a 11 c 10 g 6 t
ORIGIN

Query Match      78.7%; Score 11.8; DB 9; Length 34;
Best Local Similarity 86.7%; Pred. No. 1.3e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taccagggtgagtc 15
Db 16 TACCCAGCTGTGCT 30

RESULT 4
AR075823/c
LOCUS      AR075823      48 bp      DNA
DEFINITION Sequence 6 from patent US 5958700.
ACCESSION  AR075823
VERSION     AR075823.1 GI:10002569
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 48)
AUTHORS    Nadeau,J.G., Pitner,J.Bruce, Linn,C.Preston and Schram,J.L.
TITLE      Detection of nucleic acids by fluorescence quenching
JOURNAL    Patent: US 5958700-A 6 28-SEP-1999;
```

```

FEATURES          Location/Qualifiers
  source
  1..48
BASE COUNT      15 a      8 c      10 g      15 t
ORIGIN

Query Match
Best Local Similarity 78.7%; Score 11.8; DB 9; Length 48;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taccaggtgagtct 15
   ||| ||| ||| ||| |||
Db 41 TACTCAGATGAGTCT 27

RESULT 5
E30456/c
LOCUS      E30456      48 bp      DNA      PAT      07-FEB-2001
DEFINITION Method for detecting target nucleic acid sequence and
            oligonucleotide.
ACCESSION  E30456
VERSION     E30456.1 GI:13025613
KEYWORDS   JP 1999056380-A/6.
SOURCE     unclassified.
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 48)
AUTHORS   James,G.N.J.J., Pitona,C.P.R.R. and L.S.
TITLE     Method for detecting target nucleic acid sequence and
JOURNAL   Patent: JP 1999056380-A 6 02-MAR-1999;
          BECTON DICKINSON & CO
COMMENT   OS Unidentified
          PN JP 1999056380-A/6
          PD 02-MAR-1999
          PF 29-MAY-1998 JP 1998166141
          PR 30-MAY-1997 US 08/865,675
          PI JAMES G NADEAU,J BLUCE PITONA,C PRESTON RIN,JAMES L SHURAMU PC
          C12N15/09,C12Q1/68,G01N33/50,G01N33/566,C12N15/00 CC
Strandedness: Single;
CC Topology: Linear;
FH Key
FT source      Location/Qualifiers
          1..48
          /organism="Unidentified".
FEATURES          Location/Qualifiers
  source
  1..48
BASE COUNT      15 a      8 c      10 g      15 t
ORIGIN

Query Match
Best Local Similarity 78.7%; Score 11.8; DB 10; Length 48;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taccaggtgagtct 15
   ||| ||| ||| ||| |||
Db 41 TACTCAGATGAGTCT 27

RESULT 6
A34802/c
LOCUS      A34802      18 bp      DNA      PAT      16-JUL-1996
DEFINITION HSV probe.
ACCESSION  A34802
VERSION     A34802.1 GI:1568283
KEYWORDS   .
SOURCE     synthetic construct.
ORGANISM   synthetic construct.
REFERENCE  1 (bases 1 to 18)
AUTHORS   Renard,A. and Thiry,M.

FEATURES          Location/Qualifiers
  source
  1..48
BASE COUNT      15 a      8 c      10 g      15 t
ORIGIN

Query Match
Best Local Similarity 76.0%; Score 11.4; DB 10; Length 30;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cccaggtgagtct 15
   ||| ||| ||| |||
Db 6 CCCAGGTAAGTCT 18

RESULT 8
AR060552
LOCUS      AR060552      20 bp      DNA      PAT      29-SEP-1999
DEFINITION Sequence 53 from patent US 5840693.
ACCESSION  AR060552
VERSION     AR060552.1 GI:5987002
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 20)
AUTHORS   Eriksson,U., Olofsson,B., Alitalo,K. and Pajusola,K.
TITLE     Vascular endothelial growth factor-B
JOURNAL   Patent: US 5840693-A 53 24-NOV-1998;
          Location/Qualifiers
FEATURES          Location/Qualifiers
  source
  1..20
BASE COUNT      5 a      4 c      6 g      5 t
ORIGIN

Recombinant polypeptides of the haemorrhagic septicaemia virus in
fish
Patent: EP 0377349-A 20 11-JUL-1990;
EUROGENTEC S.A
FEATURES          Location/Qualifiers
  source
  1..18
BASE COUNT      5 a      4 c      7 g      2 t
ORIGIN

Query Match
Best Local Similarity 92.3%; Score 11.4; DB 9; Length 18;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cccaggtgagtct 15
   ||| ||| ||| |||
Db 14 CCCAGGTCAGTCT 2

RESULT 7
I36150
LOCUS      I36150      30 bp      DNA      PAT      13-MAY-1997
DEFINITION Sequence 34 from patent US 5604131.
ACCESSION  I36150
VERSION     I36150.1 GI:2087374
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 30)
AUTHORS   Wadsworth,S., Snyder,B., Reddy,V.B. and Wei,C.
TITLE     cDNA-genomic DNA hybrid sequence encoding APP770 containing a
          genomic DNA insert of the KI and OX-2 regions
JOURNAL   Patent: US 5604131-A 34 18-FEB-1997;
          Location/Qualifiers
FEATURES          Location/Qualifiers
  source
  1..30
BASE COUNT      7 a      9 c      5 g      9 t
ORIGIN

Query Match
Best Local Similarity 92.3%; Score 11.4; DB 10; Length 30;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cccaggtgagtct 15
   ||| ||| ||| |||
Db 6 CCCAGGTAAGTCT 18

```

```

Query Match      73.3%; Score 11; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.7e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 cagggtgagctc 15
Db 8 CAGGTGAGTCT 18

RESULT 9
LOCUS AX069241 20 bp DNA PAT 25-JAN-2001
DEFINITION Sequence 52 from Patent WO0102594.
ACCESSION AX069241
VERSION AX069241.1 GI:12579122
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 20)
AUTHORS Graves,D.R., Thomsen,L., Catchpole,I.R. and Ford,M.J.
TITLE Dna constructs based on the eif4a gene promoter
JOURNAL Patent: WO 0102594-A 52 11-JAN-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
Location/Qualifiers
1..20
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 6 a 6 c 7 g 1 t
ORIGIN

Query Match      73.3%; Score 11; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.7e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 acccaggtgag 12
Db 5 ACCCAGGTGAG 15

RESULT 10
LOCUS AR003365 28 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 4 from patent US 5744304.
ACCESSION AR003365
VERSION AR003365.1 GI:3964624
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE
1 (bases 1 to 28)
AUTHORS Munford,R.S.
TITLE Inflammation-induced expression of a recombinant gene
JOURNAL Patent: US 5744304-A 4 28-APR-1998;
FEATURES
Location/Qualifiers
1..28
/organism="unknown"
BASE COUNT 4 a 6 c 11 g 7 t
ORIGIN

Query Match      73.3%; Score 11; DB 9; Length 28;
Best Local Similarity 100.0%; Pred. No. 4.5e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cccaggtgagt 13
Db 16 CCCAGGTGAGT 26

RESULT 11
LOCUS AR069241 20 bp DNA PAT 25-JAN-2001
DEFINITION Sequence 52 from Patent WO0102594.
ACCESSION AX069241
VERSION AX069241.1 GI:12579122
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 20)
AUTHORS Graves,D.R., Thomsen,L., Catchpole,I.R. and Ford,M.J.
TITLE Dna constructs based on the eif4a gene promoter
JOURNAL Patent: WO 0102594-A 52 11-JAN-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
Location/Qualifiers
1..20
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 6 a 6 c 7 g 1 t
ORIGIN

Query Match      73.3%; Score 11; DB 9; Length 28;
Best Local Similarity 100.0%; Pred. No. 4.5e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cccaggtgagt 13
Db 16 CCCAGGTGAGT 26

RESULT 12
LOCUS AR043988 21 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 78 from patent US 5817311.
ACCESSION AR043988
VERSION AR043988.1 GI:5965453
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE
1 (bases 1 to 21)
AUTHORS Bazin,H. and Latinne,D.
TITLE Methods of inhibiting T-cell mediated immune responses with
LO-CD2a-specific antibodies
JOURNAL Patent: US 5817311-A 78 06-OCT-1998;
FEATURES
Location/Qualifiers
1..21
/organism="unknown"
BASE COUNT 5 a 5 c 8 g 3 t
ORIGIN

Query Match      72.0%; Score 10.8; DB 9; Length 21;
Best Local Similarity 85.7%; Pred. No. 6.4e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taccaggtgagtc 14
Db 15 TCCTCAGGTGAGTC 2

RESULT 13
LOCUS AR043989 21 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 79 from patent US 5817311.
ACCESSION AR043989
VERSION AR043989.1 GI:5965454
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE
1 (bases 1 to 21)
AUTHORS Bazin,H. and Latinne,D.
TITLE Methods of inhibiting T-cell mediated immune responses with

```

JOURNAL LO-CD2a-specific antibodies  
Patent: US 5817311-A 79 06-OCT-1998;  
FEATURES Location/Qualifiers  
source 1..21  
BASE COUNT 3 a 8 c 5 g 5 t  
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 21;  
Best Local Similarity 85.7%; Pred. No. 6.4e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taccacaggtgagtc 14  
| | | | | | | | | |  
Db 7 TCCTCAGGTGAGTC 20

RESULT 14  
AR073521/c  
LOCUS 21 bp DNA PAT 28-AUG-2000  
DEFINITION Sequence 78 from patent US 5951983.  
ACCESSION AR073521  
VERSION AR073521.1 GI:10000285  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Bazin,H., Latinne,D., Kaplan,R., Kieber-Emmons,T., Postema,C.E. and White-Scharf,M.E.  
TITLE Methods of inhibiting T cell mediated immune responses with humanized LO-CD2A-specific antibodies  
JOURNAL Patent: US 5951983-A 78 14-SEP-1999;  
FEATURES Location/Qualifiers  
source 1..21  
BASE COUNT 5 a 5 c 8 g 3 t  
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 21;  
Best Local Similarity 85.7%; Pred. No. 6.4e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taccacaggtgagtc 14  
| | | | | | | | | |  
Db 15 TCCTCAGGTGAGTC 2

RESULT 15  
AR073522  
LOCUS 21 bp DNA PAT 28-AUG-2000  
DEFINITION Sequence 79 from patent US 5951983.  
ACCESSION AR073522  
VERSION AR073522.1 GI:10000286  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Bazin,H., Latinne,D., Kaplan,R., Kieber-Emmons,T., Postema,C.E. and White-Scharf,M.E.  
TITLE Methods of inhibiting T cell mediated immune responses with humanized LO-CD2A-specific antibodies  
JOURNAL Patent: US 5951983-A 79 14-SEP-1999;  
FEATURES Location/Qualifiers  
source 1..21  
BASE COUNT 3 a 8 c 5 g 5 t  
ORIGIN

Query Match 72.0%; Score 10.8; DB 9; Length 21;  
Best Local Similarity 85.7%; Pred. No. 6.4e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taccacaggtgagtc 14  
| | | | | | | | | |  
Db 7 TCCTCAGGTGAGTC 20

Search completed: October 2, 2001, 15:56:54  
Job time: 14178 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:12 ; Search time 10798.2 Seconds  
(without alignments)  
13.131 Million cell updates/sec

Title: US-09-757-100B-31

Perfect score: 15

Sequence: 1 taccaggtgagctct 15

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 76260

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
JOURNAL plasmid inserts  
COMMENT Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0347 row: C column: 11  
Seq primer: CGTGTAAACGACGCGCAGT  
Class: plasmid ends  
High quality sequence stop: 25.  
Location/Qualifiers  
1. .25  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0347C11"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

FEATURES  
source  
1. .25  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0347C11"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 3 a 6 c 9 g 7 t  
ORIGIN  
Query Match 76.0%; Score 11.4; DB 245; Length 25;  
Best Local Similarity 92.3%; Pred. No. 3e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cccagggtgagctc 15  
Db 1 |||||  
5 CTCAGGTGAGTCT 17

RESULT 2  
AZ303920/c  
LOCUS AZ303920 32 bp DNA 29-SEP-2000  
DEFINITION LM0003F19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0003F19 R, DNA sequence.  
ACCESSION AZ303920  
VERSION AZ303920.1 GI:10339372  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 32)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	11.4	76.0	25	245	AZ506197	AZ506197 LM0347C11
C 2	11	73.3	32	241	AZ303920	AZ303920 LM0003F19
C 3	10.8	72.0	27	244	AZ495501	AZ495501 LM0331P21
C 4	10.8	72.0	43	249	AZ785625	AZ785625 LM0029L08
C 5	10.4	69.3	33	242	AZ382781	AZ382781 LM0140N08
C 6	10.4	69.3	34	103	A1873935	A1873935 wm44d08.x
C 7	10.4	69.3	35	243	AZ438946	AZ438946 LM0229L15
C 8	10.4	69.3	42	242	AZ377696	AZ377696 LM0132I11
C 9	10.4	69.3	46	190	W47560	W47560 zc35a01.r1
C 10	10.4	69.3	47	249	AZ795746	AZ795746 LM0051F07
C 11	10.4	69.3	48	250	AZ819513	AZ819513 LM0091L11
C 12	10.4	69.3	49	17	A1188352	A1188352 qd08R06.x
C 13	10.4	69.3	50	107	AU105506	AU105506 AU105506
C 14	10.4	69.3	50	107	AU105507	AU105507 AU105507
C 15	10.2	68.0	24	241	AZ303688	AZ303688 LM0003M11
C 16	10.2	68.0	43	244	AZ480568	AZ480568 LM0302M10
C 17	10.2	68.0	47	156	D38689	D38689 HUMC1264 Hu
C 18	10.2	68.0	48	258	AL462744	AL462744 T. brucei
C 19	10.2	68.0	49	168	BF740213	BF740213 hu57b11.x
C 20	10.2	68.0	50	15	A1024893	A1024893 ou52b04.x
C 21	10.2	68.0	50	107	AU106341	AU106341 AU106341
C 22	10.6	66.7	26	250	AZ827167	AZ827167 LM0103N22
C 23	10.6	66.7	34	250	AZ827876	AZ827876 LM0104M01
C 24	10.6	66.7	41	246	AZ580746	AZ580746 LM0369I11
C 25	10.6	66.7	48	188	T17575	T17575 mps v60 The
C 26	9.8	65.3	28	241	AZ313467	AZ313467 LM0029P06
C 27	9.8	65.3	28	242	AZ345677	AZ345677 LM0080B01
C 28	9.8	65.3	34	242	AZ349201	AZ349201 LM0086B08
C 29	9.8	65.3	37	102	A1862676	A1862676 wJ27f04.x
C 30	9.8	65.3	40	9	AA605733	AA605733 fai18d03.s
C 31	9.8	65.3	42	243	AZ427649	AZ427649 LM0209A15
C 32	9.8	65.3	43	244	AZ495842	AZ495842 LM0331L23
C 33	9.8	65.3	44	13	AA930614	AA930614 vv67a03.r
C 34	9.8	65.3	45	250	AZ28142	AZ28142 LM0104P24
C 35	9.8	65.3	46	23	A1663836	A1663836 uj06a03.x
C 36	9.8	65.3	47	244	AZ476231	AZ476231 LM0294P14
C 37	9.8	65.3	49	241	AZ328192	AZ328192 LM0051A19
C 38	9.8	65.3	49	246	AZ607150	AZ607150 LM0429M24
C 39	9.8	65.3	50	107	AU102408	AU102408 AU102408
C 40	9.8	65.3	50	107	AU102658	AU102658 AU102658
C 41	9.8	65.3	50	107	AU106203	AU106203 AU106203
C 42	9.8	65.3	50	188	R88739	R88739 yp93c04.s1
C 43	9.4	62.7	23	241	AZ311201	AZ311201 LM0026F04
C 44	9.4	62.7	24	246	AZ588336	AZ588336 LM0396O06
C 45	9.4	62.7	25	241	AZ330576	AZ330576 LM0056E05

ALIGNMENTS

RESULT 1  
LOCUS AZ506197 25 bp DNA GSS 05-OCT-2000  
DEFINITION LM0347C11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0347C11 F, DNA sequence.  
ACCESSION AZ506197  
VERSION AZ506197.1 GI:10687513  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 25)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

TITLE  
Journal  
COMMENT  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0003 row: F column: 19  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 32.  
Location/Qualifiers  
1. 32

## FEATURES

source

```

1. 32
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUC1M0003F19"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
5 a 10 c 9 g 8 t

```

BASE COUNT  
ORIGIN

Query Match 73.3%; Score 11; DB 241; Length 32;  
Best Local Similarity 100.0%; Pred. No. 5.3e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 acccaggtgag 12  
|||||  
Db 28 ACCCAGGTGAG 18

RESULT 3

AZ495501/c  
LOCUS  
DEFINITION  
IM0331P21F Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUC1M0331P21 F, DNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. (bases 1 to 27)  
REFERENCE  
AUTHORS  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0331 row: P column: 21  
Seq primer: CGTGTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 27.  
Location/Qualifiers  
1. 27

## FEATURES

source

```

1. 27
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUC1M0331P21"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
11 a 3 c 8 g 5 t

```

BASE COUNT  
ORIGIN

Query Match 72.0%; Score 10.8; DB 244; Length 27;  
Best Local Similarity 85.7%; Pred. No. 6.8e+04;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 taccacaggtgagtc 14  
|||||  
Db 25 TACCACATGTGATTC 12

RESULT 4

AZ785625/c  
LOCUS  
DEFINITION  
2M0029L08R Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUC2M0029L08 R, DNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. (bases 1 to 43)  
REFERENCE  
AUTHORS  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: gdunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0029 row: L column: 08  
 Seq primer: CACACAGGACACGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 43.  
 Location/Qualifiers

## FEATURES

source

```

1..43
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0029L08"
/clone_lib="Mouse 10kb plasmid-UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain:XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
6 a 20 c 10 g 7 t

```

BASE COUNT

ORIGIN

```

Query Match      72.08; Score 10.8; DB 249; Length 43;
Best Local Similarity 85.7%; Pred. No. 7e+04; Indels 0; Gaps 0;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acccaggtgagtct 15
    | | | | | | | | | |
Db 15 AGCCAGGTGAGGCT 2

```

RESULT 5

AZ382781/c

```

LOCUS      AZ382781      33 bp      DNA      GSS      02-OCT-2000
DEFINITION  IM0140N08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0140N08 F, DNA sequence.
ACCESSION  AZ382781
VERSION    AZ382781.1 GI:10496481
KEYWORDS   GSS.
SOURCE     house musculus
ORGANISM   Mus musculus

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 33)

```

## AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: gdunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0140 row: N column: 08  
 Seq primer: CGTTGTAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 33.  
 Location/Qualifiers

## TITLE

## JOURNAL

## COMMENT

## FEATURES

source

```

1..33
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0140N08"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
4 a 10 c 8 g 11 t

```

BASE COUNT

ORIGIN

```

Query Match      69.3%; Score 10.4; DB 242; Length 33;
Best Local Similarity 91.7%; Pred. No. 1.2e+05;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 taccacaggtgag 12
    | | | | | | | | | |
Db 20 TACCACAGGTGCG 9

```

RESULT 6

AI873935/c

LOCUS

```

DEFINITION  AI873935      34 bp      mRNA      EST      01-SEP-1999
            Wm44d08.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2438799 3'
            similar to TR:070396 070396 SIK SIMILAR PROTEIN. ; mRNA sequence.
ACCESSION  AI873935
VERSION    AI873935.1 GI:5547984
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 34)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 Clone Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality

Seq primer: -40UP from Gibco

High quality sequence stop: 1.

## FEATURES

source

1..34

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2438799"

/clone\_lib="NCI-CGAP\_Ut4"

/tissue\_type="serous papillary carcinoma, high grade, 2

pooled tumors"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.48 kb. Life Technologies catalog #:

11542-016"

7 a 8 c 7 g 12 t

BASE COUNT

ORIGIN

Query Match 69.3%; Score 10.4; DB 103; Length 34;  
 Best Local Similarity 91.7%; Pred. NO. 1.2e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 acccaggtgagt 13

|||||

Db 34 ACACAGGTGAGT 23

RESULT 7

AZ438946

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Robert B. Weiss

University of Utah

Rm. 308, Biomedical

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

20 S. 2030 E., SLC, UT

100b-31.szlim50.rst

Insert Length: 10000 Std Error: 0.00

Plate: 0229 row: L column: 15

Seq primer: CGTTGTAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 35.

## FEATURES

source

1..35

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0229L15"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, P-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

9 a 9 c 8 g 9 t

BASE COUNT

ORIGIN

Query Match 69.3%; Score 10.4; DB 243; Length 35;

Best Local Similarity 91.7%; Pred. NO. 1.2e+05;

Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 taccacaggtgag 12

|||||

Db 10 TACACAGGTGAG 21

RESULT 8

AZ377696

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Robert B. Weiss

University of Utah

Rm. 308, Biomedical

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

20 S. 2030 E., SLC, UT

100b-31.szlim50.rst

Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0132 row: 1 column: 11  
 Seq primer: CGTTGTAACAGCGGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 42.  
 Location/Qualifiers

# FEATURES

1. .42  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0132111"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

# BASE COUNT

12 a 12 c 9 g 9 t

Query Match 69.3%; Score 10.4; DB 242; Length 42;  
 Best Local Similarity 91.7%; Pred. No. 1.2e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ccagggtgagtc 14  
 |||||  
 Db 2 CCCATGTGAGTC 13

# RESULT 9

W47560 46 bp mRNA EST 11-OCT-1996  
 zc35a01.r1 Soares.senescent.fibroblasts\_NbHSF Homo sapiens cDNA  
 Clone IMAGE:324264 5', similar to PIR:S43417 S43417 RNA/DNA-binding  
 Protein - mouse ;, mRNA sequence.

W47560  
 EST  
 W47560.1 GI:1332239  
 human.

# ORGANISM

Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

# REFERENCE

AUTHORS  
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman  
 .M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,  
 Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston  
 ,R., Williamson,A., Wohlmann,P. and Wilson,R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

# TITLE

JOURNAL  
 COMMENT

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Insert Length: 2909 Std Error: 0.00  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 1.  
 Location/Qualifiers

# FEATURES

1. .46  
 /organism="Homo sapiens"  
 /db\_xref="GDB:125577g"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:324264"  
 /clone\_lib="Soares.senescent.fibroblasts\_NbHSF"  
 /tissue\_type="senescent fibroblast"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Vector: pF7T3D (Pharmacia) with a modified  
 polylinker V-TYPE: phagemid; Site\_1: Not I; Site\_2: Eco RI  
 ; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer [5',  
 TGTACCAATCTGAAGTGGAGCGCGGCATTTTTTTTTTTTTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pF7T3 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo."

# BASE COUNT

16 a 13 c 10 g 6 t 1 others

Query Match 69.3%; Score 10.4; DB 190; Length 46;  
 Best Local Similarity 91.7%; Pred. No. 1.2e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ccagggtgagtc 15  
 |||||  
 Db 36 CCGGGTGAAGTCT 25

# RESULT 10

AZ795746  
 LOCUS  
 DEFINITION  
 2M051F07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC2M051F07 F, DNA sequence.

ACCESSION  
 AZ795746  
 VERSION  
 AZ795746.1 GI:12943089  
 GSS.  
 KEYWORDS  
 house mouse.

# ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 47)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 .M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu

# TITLE

JOURNAL  
 COMMENT

Insert Length: 10000 Std Error: 0.00  
 Plate: 0051 row: F column: 07  
 Seq primer: CGTTGTAACAGCGGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 47.

FEATURES  
source

Location/Qualifiers  
1. 47  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="UUGC2M0051F07"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 6 a 17 c 14 g 10 t  
ORIGIN

Query Match 69.3%; Score 10.4; DB 249; Length 47;  
Best Local Similarity 91.7%; Pred. No. 1.2e+05;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ccaggtgagtc 14  
||||| |||||  
Db 24 CCCAAGTGAGTC 35

## RESULT 11

A2819513 48 bp DNA GSS 20-FEB-2001  
LOCUS  
DEFINITION 2M009111F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0091L11 F, DNA sequence.  
ACCESSION A2819513  
VERSION A2819513.1 GI:12989421  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 48)  
REFERENCE Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0091 row: L column: 11  
Seq primer: CGTGTAAACGACGCCAGT  
Class: plasmid ends

## FEATURES

## source

High quality sequence stop: 48.

Location/Qualifiers  
1. 48  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="UUGC2M0091L11"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 10 c 18 g 15 t  
ORIGIN

Query Match 69.3%; Score 10.4; DB 250; Length 48;  
Best Local Similarity 91.7%; Pred. No. 1.2e+05;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ccaggtgagtc 15  
||||| |||||  
Db 35 CCAGCTGAGTCT 46

## RESULT 12

A1188352 49 bp mRNA EST 28-OCT-1998  
LOCUS  
DEFINITION qd08b06.xl Soares.placenta.8to9weeks.2NbHP8to9w Homo sapiens cDNA clone IMAGE:1723091 3' similar to TR:Q95341 Q95341 UBIQUINOL  
CYTOCHROME C REDUCTASE ;, mRNA sequence.  
ACCESSION A1188352  
VERSION A1188352.1 GI:3739561  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 49)  
REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
TITLE Unpublished (1997)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgaps-r@mail.nih.gov  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Trace considered overall poor quality  
Insert Length: 1262 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 49  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1723091"



/clone\_lib="Soares\_placenta\_8to9weeks\_2NbHP8to9w"  
 /dev\_stage="two placentae: one from 8 weeks and another  
 from 9 weeks post conception"  
 /lab\_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT7T3D.(Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACCAATCTCAAGTCGAGCGCGGATTTTGTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT7T3 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M.Fatima Bonaldo."

BASE COUNT 15 a 14 c 8 g 12 t  
 ORIGIN

Query Match 69.3%; Score 10.4; DB 17; Length 49;  
 Best Local Similarity 91.7%; Pred. No. 1.2e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cccaggtgagtc 14  
 | | | | | | | | | |  
 Db 3 CCCAGATGAGTC 14

## RESULT 13

AU105506 50 bp mRNA EST 05-APR-2001  
 LOCUS AU105506 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 DEFINITION HRC12806, mRNA sequence.

ACCESSION AU105506  
 VERSION AU105506.1 GI:13555027  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.  
 TITLE Fine structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: yusuzuki@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

source  
 1. .50  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HRC12806"

/clone\_lib="Sugano Homo sapiens cDNA library"  
 BASE COUNT 11 a 14 c 16 g 9 t  
 ORIGIN

Query Match 69.3%; Score 10.4; DB 107; Length 50;  
 Best Local Similarity 91.7%; Pred. No. 1.2e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 taccaggtgag 12  
 | | | | | | | | | |  
 Db 31 TCCCCAGGTGAG 42

## RESULT 14

## AU105507

## LOCUS

AU105507 50 bp mRNA EST 05-APR-2001  
 DEFINITION AU105507 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 HRC12884, mRNA sequence.

## ACCESSION

AU105507  
 VERSION AU105507.1 GI:13555028  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.

## TITLE

Fine structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries

## JOURNAL

Unpublished (2001)

## COMMENT

Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: yusuzuki@ims.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

## FEATURES

source  
 1. .50  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HRC12884"

/clone\_lib="Sugano Homo sapiens cDNA library"  
 BASE COUNT 11 a 14 c 17 g 8 t  
 ORIGIN

Query Match 69.3%; Score 10.4; DB 107; Length 50;  
 Best Local Similarity 91.7%; Pred. No. 1.2e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 taccaggtgag 12  
 | | | | | | | | | |  
 Db 31 TCCCCAGGTGAG 42

## RESULT 15

## LOCUS

AZ303688 24 bp DNA GSS 29-SEP-2000  
 DEFINITION 1M0003M11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0003M11 F, DNA sequence.

## ACCESSION

AZ303688  
 VERSION AZ303688.1 GI:10339066  
 KEYWORDS GSS.  
 SOURCE house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

## REFERENCE

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Relly  
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D. Weiss,R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Robert B. Weiss  
 University of Utah Genome Center

University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00  
Plate: 0003 row: M column: 11  
Seq primer: CGTTGTAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 24.

FEATURES

Location/Qualifiers

1..24  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGC1M0003M11"  
/clone\_lib="Mouse 10kb plasmid UGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 7 a 3 c 4 g 10 t  
ORIGIN

Query Match 68.0%; Score 10.2; DB 241; Length 24;  
Best Local Similarity 80.0%; Pred. No. 1.5e+05;  
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 taccagggtgagct 15  
||| ||| | |||  
Db 17 TATCCAGATAAGTCT 3

Search completed: October 2, 2001, 15:01:13  
Job time: 10842 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:56:54 ; Search time 3339.34 Seconds  
(without alignments)  
69.480 Million cell updates/sec

Title: US-09-757-100B-33  
Perfect score: 15  
Sequence: 1 tggcttatcttcagt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 423528  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pil.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rol2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	12.4	82.7	27	9 AR042372	AR042372 Sequence
c 2	12.4	82.7	27	9 AR052280	AR052280 Sequence
c 3	12	80.0	21	10 AX095001	AX095001 Sequence
c 4	11.8	78.7	21	9 AX002270	AX002270 Sequence
c 5	11.8	78.7	21	10 E28251	E28251 Purificatio
c 6	11.8	78.7	30	9 AR012670	AR012670 Sequence
c 7	11.8	78.7	30	10 I25036	I25036 Sequence
c 8	11.8	78.7	30	10 I30498	I30498 Sequence

9 11.8 78.7 30 10 172560 Sequence 9  
10 11.4 76.0 21 10 124623 Sequence 7  
11 11.4 76.0 21 10 125220 Sequence 7  
c 12 11 73.3 21 10 AX093520  
13 11 73.3 24 9 AX040915 Sequence  
14 11 73.3 24 9 AR051822 Sequence  
15 11 73.3 24 10 I36212  
16 11 73.3 31 9 A27527  
c 17 11 73.3 40 9 A65367 Sequence 3  
18 11 73.3 41 9 AR040914 Sequence  
c 19 11 73.3 41 9 AR051821  
c 20 11 73.3 41 10 I36211  
21 10.8 72.0 16 9 AR074822  
22 10.8 72.0 17 9 AR035054  
23 10.8 72.0 17 10 I16172  
c 24 10.8 72.0 20 9 A42061  
25 10.8 72.0 20 9 AR031048  
26 10.8 72.0 20 9 AR092977  
27 10.8 72.0 20 54 HUM06481A  
c 28 10.8 72.0 21 10 E25007  
c 29 10.8 72.0 24 94 MMTG3C6X1  
c 30 10.8 72.0 24 94 MMTCPALAC  
c 31 10.8 72.0 30 9 AR001359  
c 32 10.8 72.0 30 9 AR044844  
c 33 10.8 72.0 30 9 AR044854  
c 34 10.8 72.0 30 9 AR052230  
c 35 10.8 72.0 30 9 AR052240  
c 36 10.8 72.0 30 9 AR078339  
c 37 10.8 72.0 30 9 AR085190  
c 38 10.8 72.0 30 10 I33894  
c 39 10.8 72.0 30 10 I33904  
c 40 10.8 72.0 31 9 AR090061  
c 41 10.8 72.0 31 9 AX009745  
c 42 10.8 72.0 42 9 AR079097  
c 43 10.8 72.0 42 10 I34798  
c 44 10.8 72.0 42 10 I64516  
c 45 10.8 72.0 50 9 A23256

## ALIGNMENTS

RESULT 1  
AR042372/c  
LOCUS AR042372 27 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 11 from patent US 5811304.  
ACCESSION AR042372  
VERSION AR042372.1 GI:5962868  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 27)  
AUTHORS Huang, S.  
TITLE Nucleic acid molecules encoding retinoblastoma protein-interacting  
zinc finger proteins  
JOURNAL Patent: US 5811304-A 11 22-SEP-1998;  
FEATURES Location/Qualifiers  
source 1..27  
BASE COUNT 9 a 7 c 9 g 2 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 9; Length 27;  
Best Local Similarity 92.9%; Pred. No. 1.5e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttcag 14  
||||| |||||  
Db 27 TGGCTTTCTTCAG 14

RESULT 2  
AR052280/c  
LOCUS AR052280 27 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 11 from patent US 5811008.  
ACCESSION AR052280  
VERSION AR052280.1 GI:5975644  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 27)  
AUTHORS Huang, S.  
TITLE Retinoblastoma protein-interacting zinc finger proteins  
JOURNAL Patent: US 5811008-A 11 03-NOV-1998;  
FEATURES Location/Qualifiers  
source 1..27  
BASE COUNT 9 a 7 c 9 g 2 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 9; Length 27;  
Best Local Similarity 92.9%; Pred. No. 1.5e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttcag 14  
||||| |||||  
Db 27 TGGCTTTCTTCAG 14

RESULT 3  
AX095001  
LOCUS AX095001 21 bp DNA PAT 30-MAR-2001  
DEFINITION Sequence 179 from Patent WO0118250.  
ACCESSION AX095001  
VERSION AX095001.1 GI:13511204  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Lander, E.S.; Gargill, M., Ireland, J.S., Bolck, S., Daley, G.G. and  
McCarthy, J.J.  
TITLE Single nucleotide polymorphisms in genes  
Patent: WO 0118250-A 179 15-MAR-2001;  
JOURNAL WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH (US) ; Millennium  
Pharmaceuticals, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..21  
BASE COUNT 3 a 6 c 4 g 7 t 1 others  
ORIGIN

Query Match 80.0%; Score 12; DB 10; Length 21;  
Best Local Similarity 85.7%; Pred. No. 2.6e+04;  
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttcag 14  
||||| |||||  
Db 4 TGGCTGAYCTTCAG 17

RESULT 4  
AX002270/c  
LOCUS AX002270 21 bp DNA PAT 10-MAR-2000  
DEFINITION Sequence 7 from Patent EP0881288.  
ACCESSION AX002270  
VERSION AX002270.1 GI:7241961  
KEYWORDS  
SOURCE unidentified.

```

ORGANISM      unidentified
REFERENCE      unclassified.
1 (bases 1 to 21)
AUTHORS      Berglund,E.D. and Kirschbaum,B.D.
TITLE        Purification of higher order transcription complexes from
              transgenic non-human animals
JOURNAL      Patent: EP 0881288-A 7 02-DEC-1998;
              HOECHST AG (DE)
FEATURES      Location/Qualifiers
              source
                1..21
                /organism="unidentified"
                /db_xref="taxon:32644"
              exon
                6 a      4 c      8 g      3 t
BASE COUNT    6 a      4 c      8 g      3 t
ORIGIN
Query Match   78.7%; Score 11.8; DB 9; Length 21;
Best Local Similarity 86.7%; Pred. No. 3.4e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15
||||| ||| ||||| |||
Db 19 TGGCCTAACTTCAGT 5

RESULT 5
E28251/c
LOCUS      E28251      21 bp      DNA      PAT      07-FEB-2001
DEFINITION Purification of higher transcription complex from nonhuman
              transgenic animal.
ACCESSION  E28251
VERSION     E28251.1 GI:13025285
KEYWORDS    JP 1999004638-A/3.
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE    1 (bases 1 to 21)
AUTHORS      Bernd,K.E.B. and Meisterernst,G.P.
TITLE        Purification of higher transcription complex from nonhuman
              transgenic animal
JOURNAL      Patent: JP 1999004638-A 3 12-JAN-1999;
              HOECHST AG
COMMENT      OS Unidentified
              PN JP 1999004638-A/3
              PD 12-JAN-1999
              PF 26-MAY-1998 JP 1998144743
              PR 26-MAY-1997 DE
              PI BERND KIRSCHBAUM,ERIC BERUGURUNTO,MICHAEL MEISTERERNST, PI
              GREG PORITTSU
              PC A01K67/027,C12N15/09,C12P21/02//C12P21/08,(C12N15/09,
              C12R1:91), PC C12N15/00,
              PC (C12N15/00,C12R1:91)
              CC Strandedness: Single;
              CC Topology: Linear;
              FH Key Location/Qualifiers
              FT exon      1..21.
              /organism="unidentified"
              /db_xref="taxon:32644"
              source
                1..21
                /organism="unidentified"
                /db_xref="taxon:32644"
              BASE COUNT    6 a      4 c      8 g      3 t
ORIGIN
Query Match   78.7%; Score 11.8; DB 10; Length 21;
Best Local Similarity 86.7%; Pred. No. 3.4e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15
||||| ||| ||||| |||
Db 19 TGGCCTAACTTCAGT 5

```

```

RESULT 6
LOCUS      AR012670      30 bp      DNA      PAT      05-DEC-1998
DEFINITION Sequence 9 from patent US 5763584.
ACCESSION  AR012670
VERSION     AR012670.1 GI:3970988
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE    1 (bases 1 to 30)
AUTHORS      Godowski,P.J.
TITLE        Receptor activation with hepatocyte growth factor agonists
JOURNAL      Patent: US 5763584-A 9 09-JUN-1998;
              Location/Qualifiers
FEATURES      source
                1..30
                /organism="unknown"
                /db_xref="taxon:32644"
              BASE COUNT    7 a      7 c      4 g      12 t
ORIGIN
Query Match   78.7%; Score 11.8; DB 9; Length 30;
Best Local Similarity 86.7%; Pred. No. 3.3e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15
||||| ||| ||||| |||
Db 7 TGACATATCTTCAGT 21

RESULT 7
LOCUS      I25036      30 bp      DNA      PAT      07-OCT-1996
DEFINITION Sequence 10 from patent US 5547856.
ACCESSION  I25036
VERSION     I25036.1 GI:1604906
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE    1 (bases 1 to 30)
AUTHORS      Godowski,P.J., Lokker,N.A. and Mark,M.R.
TITLE        Hepatocyte growth factor variants
JOURNAL      Patent: US 5547856-A 10 20-AUG-1996;
              Location/Qualifiers
FEATURES      source
                1..30
                /organism="unknown"
                /db_xref="taxon:32644"
              BASE COUNT    7 a      7 c      4 g      12 t
ORIGIN
Query Match   78.7%; Score 11.8; DB 10; Length 30;
Best Local Similarity 86.7%; Pred. No. 3.3e+04;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15
||||| ||| ||||| |||
Db 7 TGACATATCTTCAGT 21

RESULT 8
LOCUS      I30498      30 bp      DNA      PAT      06-FEB-1997
DEFINITION Sequence 10 from patent US 5580963.
ACCESSION  I30498
VERSION     I30498.1 GI:1821289
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE    1 (bases 1 to 30)
AUTHORS      Godowski,P.J., Lokker,N.A. and Mark,M.R.

```

TITLE Single-chain hepatocyte growth factor variants  
JOURNAL Patent: US 5580963-A 10 03-DEC-1996;  
FEATURES Location/Qualifiers  
Source 1..30  
BASE COUNT 7 a 7 c 4 g 12 t  
ORIGIN

Query Match 78.7%; Score 11.8; DB 10; Length 30;  
Best Local Similarity 86.7%; Pred. No. 3.3e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggcttatcttcagt 15  
||| ||||| |||||  
Db 7 TGACATATCTTCAGT 21

RESULT 9

LOCUS I72560 30 bp DNA PAT 03-APR-1998  
DEFINITION Sequence 9 from patent US 5684136.  
ACCESSION I72560  
VERSION I72560.1 GI:3008699  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 30)  
AUTHORS Godowski, P.J.

TITLE Chimeric hepatocyte growth factor (HGF) ligand variants  
JOURNAL Patent: US 5684136-A 9 04-NOV-1997;  
FEATURES Location/Qualifiers  
Source 1..30  
BASE COUNT 7 a 7 c 4 g 12 t  
ORIGIN

Query Match 78.7%; Score 11.8; DB 10; Length 30;  
Best Local Similarity 86.7%; Pred. No. 3.3e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggcttatcttcagt 15  
||| ||||| |||||  
Db 7 TGACATATCTTCAGT 21

RESULT 10

LOCUS I24623 21 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 7 from patent US 5545527.  
ACCESSION I24623  
VERSION I24623.1 GI:1604493  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 21)  
AUTHORS Stevens, J.K. and Dunn, J.M.

TITLE Method for testing for mutations in DNA from a patient sample  
JOURNAL Patent: US 5545527-A 7 13-AUG-1996;  
FEATURES Location/Qualifiers  
Source 1..21  
BASE COUNT 3 a 2 c 6 g 10 t  
ORIGIN

Query Match 76.0%; Score 11.4; DB 10; Length 21;  
Best Local Similarity 92.3%; Pred. No. 5.8e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 gcttatcttcagt 15  
||| ||||| |||||  
Db 2 GCTTATGTTTCAGT 14

RESULT 11

LOCUS I25220 21 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 7 from patent US 5550020.  
ACCESSION I25220  
VERSION I25220.1 GI:1605090  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 21)  
AUTHORS Gallie, B.D., Dunn, J.M. and Stevens, J.K.

TITLE Method, reagents and kit for diagnosis and targeted screening for retinoblastoma  
JOURNAL Patent: US 5550020-A 7 27-AUG-1996;  
FEATURES Location/Qualifiers  
Source 1..21  
BASE COUNT 3 a 2 c 6 g 10 t  
ORIGIN

Query Match 76.0%; Score 11.4; DB 10; Length 21;  
Best Local Similarity 92.3%; Pred. No. 5.8e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 gcttatcttcagt 15  
||| ||||| |||||  
Db 2 GCTTATGTTTCAGT 14

RESULT 12

LOCUS AX093520/c 21 bp DNA PAT 30-MAR-2001  
DEFINITION Sequence 50 from Patent WO0118198.  
ACCESSION AX093520  
VERSION AX093520.1 GI:13509959  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Weissenbach, J. and Hazan, J.

TITLE Cloning, expression and characterisation of the spg4 gene responsible for the most frequent form of autosomal spastic paraplegia  
JOURNAL Patent: WO 0118198-A 50 15-MAR-2001;  
FEATURES Location/Qualifiers  
Source 1..21  
BASE COUNT 8 a 4 c 4 g 5 t  
ORIGIN

Query Match 73.3%; Score 11; DB 10; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ttatcttcagt 15  
||| ||||| |||||  
Db 13 TTATCTTCAGT 3

RESULT 13

AR040915

LOCUS AR040915 24 bp DNA 29-SEP-1999  
DEFINITION Sequence 14 from patent US 5811238.  
ACCESSION AR040915  
VERSION AR040915.1 GI:5961411  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 24)  
AUTHORS Stemmer,W.P.C. and Cramer,A.  
TITLE Methods for generating polynucleotides having desired characteristics by iterative selection and recombination  
JOURNAL Patent: US 5811238-A 14 22-SEP-1998;  
FEATURES  
source Location/Qualifiers  
1. .24 /organism="unknown"  
BASE COUNT 5 a 6 c 1 g 12 t  
ORIGIN

Query Match 73.3%; Score 11; DB 9; Length 24;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cttatcttcag 14  
|||||  
Db 2 CTTATCTTCAG 12

RESULT 14  
LOCUS AR051822 24 bp DNA 29-SEP-1999  
DEFINITION Sequence 14 from patent US 5830721.  
ACCESSION AR051822  
VERSION AR051822.1 GI:5975186  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 24)  
AUTHORS Stemmer,W.P.C. and Cramer,A.  
TITLE DNA mutagenesis by random fragmentation and reassembly  
JOURNAL Patent: US 5830721-A 14 03-NOV-1998;  
FEATURES  
source Location/Qualifiers  
1. .24 /organism="unknown"  
BASE COUNT 5 a 6 c 1 g 12 t  
ORIGIN

Query Match 73.3%; Score 11; DB 9; Length 24;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cttatcttcag 14  
|||||  
Db 2 CTTATCTTCAG 12

RESULT 15  
LOCUS I36212 24 bp DNA 13-MAY-1997  
DEFINITION Sequence 14 from patent US 5605793.  
ACCESSION I36212  
VERSION I36212.1 GI:2086725  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 24)  
AUTHORS Stemmer,W.P.C.  
TITLE Methods for in vitro recombination  
JOURNAL Patent: US 5605793-A 14 25-FEB-1997;

FEATURES  
source Location/Qualifiers  
1. .24 /organism="unknown"  
BASE COUNT 5 a 6 c 1 g 12 t  
ORIGIN

Query Match 73.3%; Score 11; DB 10; Length 24;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cttatcttcag 14  
|||||  
Db 2 CTTATCTTCAG 12

Search completed: October 2, 2001, 15:56:55  
Job time: 14179 sec





1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est33.\*  
26: gb\_est34.\*  
27: gb\_est35.\*  
28: gb\_est36.\*  
29: gb\_est37.\*  
30: gb\_est38.\*  
31: gb\_est39.\*  
32: gb\_est40.\*  
33: em\_estba.\*  
34: em\_estfun.\*  
35: em\_esthum1.\*  
36: em\_esthum2.\*  
37: em\_esthum3.\*  
38: em\_esthum4.\*  
39: em\_esthum5.\*  
40: em\_esthum6.\*  
41: em\_esthum7.\*  
42: em\_esthum8.\*  
43: em\_esthum9.\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: dunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0034 row: M column: 09  
 Seq primer: CACACAGGAACACGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 21.  
 Location/Qualifiers  
 1. 21

FEATURES  
 source  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /notes="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gil4732114[gb]/AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 10 a 3 c 3 g 5 t  
 ORIGIN  
 Query Match 78.7% Score 11.8; DB 249; Length 21;  
 Best Local Similarity 86.7%; Pred. No. 2e+04;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15  
 ||| |||||  
 Db 18 TGGTTTATCTTCAAT 4

RESULT 2  
 AZ830358 43 bp DNA GSS 20-FEB-2001  
 LOCUS 2M0109M04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC2M0109M04 R, DNA sequence.  
 ACCESSION AZ830358  
 VERSION AZ830358.1 GI:13000266  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 43)  
 REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	11.8	78.7	21	249	AZ787920 2M0034M09
C 2	11.4	76.0	23	250	AZ830358 2M0109M04
C 3	10.8	72.0	23	247	AZ649569 1M0519M09
C 4	10.8	72.0	40	258	TA187A020
C 5	10.4	69.3	28	16	AT118907
C 6	10.4	69.3	37	13	AA906910 0J70B12.S
C 7	10.4	69.3	37	14	AA946896 0Q53F01.S
C 8	10.4	69.3	38	156	D45802 HUMG503019
C 9	10.4	69.3	42	106	AU010757 AU010757
C 10	10.4	69.3	45	249	AZ772066 1M0574L08
C 11	10.4	69.3	46	241	AZ324419 1M0046B20
C 12	10.4	69.3	46	249	AZ793395 2M0046B07
C 13	10.4	69.3	49	120	AW772664 xN02B05.Y
C 14	10.4	69.3	50	107	AU103710 AU103710
C 15	10.2	68.0	22	258	TA288E04P
C 16	10.2	68.0	37	23	AI664274
C 17	10.2	68.0	42	246	AZ593631 1M0405H14
C 18	10.2	68.0	44	13	AA883443 am25D01.S
C 19	10.2	68.0	47	172	BG030013 602297169
C 20	10.2	68.0	50	251	AZ919467 1006015E0
C 21	10.6	66.7	25	23	AI647975 UK38F04.X
C 22	10.6	66.7	36	250	AZ809373 2M0073A20
C 23	10.6	66.7	39	258	TA129B06P
C 24	10.6	66.7	40	12	AA862574
C 25	10.6	66.7	44	247	AZ660568 1M0538B15
C 26	10.6	66.7	45	18	AI309096 1B57G10.X
C 27	10.6	66.7	49	11	AA711423 vU25E09.R
C 28	9.8	65.3	27	189	T96372 ye49e01.s1
C 29	9.8	65.3	31	113	AW247904
C 30	9.8	65.3	31	250	AZ849227 2M0150N02
C 31	9.8	65.3	31	258	TA52H11Q
C 32	9.8	65.3	32	243	AZ420867 1M0198F22
C 33	9.8	65.3	36	12	AA860352 aJ59B01.S
C 34	9.8	65.3	39	249	AZ767935 1M0567E12
C 35	9.8	65.3	42	249	AZ772691
C 36	9.8	65.3	46	164	BZ209004 SMOV53117
C 37	9.8	65.3	49	258	TA264D06Q
C 38	9.8	65.3	50	104	AJ237103
C 39	9.8	65.3	50	107	AU102434 AU102434
C 40	9.8	65.3	50	107	AU102435 AU102435
C 41	9.8	65.3	50	107	AU102436 AU102436
C 42	9.8	65.3	50	107	AU102437 AU102437
C 43	9.8	65.3	50	107	AU102439 AU102439
C 44	9.8	65.3	50	107	AU102440 AU102440
C 45	9.8	65.3	50	107	AU102441 AU102441

## ALIGNMENTS

RESULT 1  
 AZ787920/c 21 bp DNA GSS 16-FEB-2001  
 LOCUS 2M0034M09R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC2M0034M09 R, DNA sequence.  
 ACCESSION AZ787920  
 VERSION AZ787920.1 GI:12927197  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 21)  
 REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.

and Wright, D. Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000).  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: [ddunne@genetics.utah.edu](mailto:ddunne@genetics.utah.edu)  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0109 row: M column: 04  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 43.  
 Location/Qualifiers

FEATURES source

FEATURES  
source

```

1. .23
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0519N09"
/clone_lib="Mouse 10kb p

```

**Vector:** pWD42nv; Purified genomic DNA from *M. musculus* C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (q14F123114[gb|AF129072.1], a copy-number inducible derivative of plasmid R1). The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

BASE COUNT	11 a	4 c	3 q	5 t

```
Query Match          76.0%; Score 11.4; DB 250; Length 43;
Best Local Similarity 92.3%; Pred. No. 3.7e+04;
Matches 12: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy	1	tggcttatcttca	13
Db	8	TGGCTTATCTTTA	20

RESULT 3  
AZ649569/C

<b>RESULT</b>	<b>3</b>				
AZ649569/c					
LOCUS					
DEFINITION	AZ649569	23 bp	DNA	GSS	14-DEC-2000
	1M0519N09F	Mouse	10kb plasmid	UUGC1M library	Mus musculus genomic
	clone UUGC1M0519N09 F,				DNA sequence.

ACCESSION	AZ649569
VERSION	AZ649569.1

KEYWORDS	GSS.
SOURCE	house mouse.

ORGANISM	Mus musculus	Eukaryota; M
muscle	muscle	

1 (bases 1 to 23)

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah  
Genome Center  
University of Utah  
Rm. 308, Biomedical  
84112, USA  
Tel.: 801 585 5606

Tel: 801 585 5606  
Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
 Insert Length: 10000    Std Error: 0.00

insert length: 1000    seq error: 0.00  
Plate: 0519    row: N    column: 09  
Seq primer: CGTTGTAAACGACGGCCAGT

seq primer: CCGTGTAAACACCGGCCAGT  
Class: plasmid ends  
High quality sequence stop: 23.

**FEATURES**

```

1. .23
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0519N09"
/clone_lib="Mouse 10kb p

```

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="vector: PWD42hv; Purified genomic DNA from M.  
musculus C57Bl/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (g1147321141gb1AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

BASE COUNT	11 a	4 c	3 q	5 t
	and selected for amplification			

```
Query Match          72.0%; Score 10.8; DB 247; Length 23;
Best Local Similarity 85.7%; Pred. No. 7.5e+04;
Matches 12: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 tggcttatcttcag 14  
 |||||  
 Db 20 TGGCATATTTTCAG 7

RESULT 4  
TA187A020/C

RESULT	4
TA187A02Q/c	
LOCUS	TA187A02Q 40 bp DNA GSS 13-DEC-2000
DEFINITION	T. brucei sheared genomic DNA clone 187a02, reverse sequence, genomic survey sequence.

Accession	Version	GI	Sequence
AL476274			
AL476274.1		GI:11841048	

**KEYWORDS**  
GSS.

**SOURCE**  
Trypanosoma brucei.

ORGANISM  
Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.  
1 (bases 1 to 40)

**AUTHORS** Hall, N., Bowman, S., Lennard, N. J., Doggett, J., Atkin, R.,

**TITLE**  
**JOURNAL**  
**COMMENT**

Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,  
 Melville,S.E., Rajandream,M.A. and Barrell,B.G.  
 Direct Submission  
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
 Project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
 nhls@sanger.ac.uk  
 Constructed at the Institute for Genomic Research (TIGR),  
 Rockville, MD. Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 CUTat 10.1) was mechanically sheared  
 to give a tight size distribution (4 kb). The v + i method used for the library construction is  
 described in detail in Smith, H. and Venter, J.C. (Making small  
 insert libraries for whole genome shotgun sequencing projects. In  
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Barrell, Oxford University Press, 1999).  
 Email: nelsayed@tigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available  
 at [http://www.sanger.ac.uk/projects/T\\_brucei/](http://www.sanger.ac.uk/projects/T_brucei/).

**FEATURES**  
 source

1. .40  
 Location/Qualifiers  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="187a02"

**BASE COUNT** 9 a 9 c 14 g 8 t

**ORIGIN**

Query Match 72.0%; Score 10.8; DB 258; Length 40;  
 Best Local Similarity 85.7%; Pred. No. 8e+04;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgcttcttcag 14

||||| |||||

Db 23 TGGGTCTCTTCAG 10

**RESULT**  
 5  
 AII18907/c

**LOCUS**

DEFINITION uc15f07.r1 Soares\_mammary\_gland\_NBMWG Mus musculus cDNA clone  
 IMAGE:1398085 5' similar to TR:Q13825 Q13825 AU-BINDING  
 PROTEIN/ENOYL-COA HYDRATASE. ; mRNA sequence.

**ACCESSION**  
 VERSION  
 KEYWORDS  
 SOURCE

AII18907 28 bp mRNA EST 02-SEP-1998  
 AII18907.1 GI:3519231  
 EST.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 28)

**REFERENCE**  
 AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

**TITLE**  
**JOURNAL**  
**COMMENT**

The WashU-HHMI Mouse EST Project  
 Unpublished (1996)  
 Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLM; contact the  
 IMAGE Consortium ([info@image.llnl.gov](http://info@image.llnl.gov)) for further information.  
 MGI:909801

Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -28m13 rev2 Et from Amersham  
 High quality sequence stop: 1.

**FEATURES**  
 source

1. .28  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1398085"  
 /clone\_lib="Soares\_mammary\_gland\_NBMWG"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"

/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia)  
 ) with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer [5'  
 TCTTACCATCTGAAGTGGAGCGCGGATGGTTTTTTTTTTTTTTTTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT7T3 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M.Fatima  
 Bonaldo."

**BASE COUNT** 9 a 4 c 10 g 5 t

**ORIGIN**

Query Match 69.3%; Score 10.4; DB 16; Length 28;  
 Best Local Similarity 91.7%; Pred. No. 1.3e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cttatcttcagt 15

||||| ||

Db 19 CTTATCTTCGT 8

**RESULT**  
 6  
 AA906910/c

**LOCUS**

DEFINITION AA906910 37 bp mRNA EST 09-JUN-1998  
 o170b12.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
 IMAGE:1503647 3' similar to TR:Q33565 Q33565 EATRO 164 KINETOPLAST  
 CR3. ; mRNA sequence.

**ACCESSION**  
 VERSION  
 KEYWORDS  
 SOURCE

AA906910  
 AA906910.1 GI:3042370  
 EST.  
 human.

**ORGANISM**

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**  
 AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/hicqap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

**JOURNAL**  
**COMMENT**

Contact: Robert Strausberg, Ph.D.  
 Email: cgaps@remail.nih.gov  
 This clone is available royalty-free through LLM; contact the  
 IMAGE Consortium ([info@image.llnl.gov](http://info@image.llnl.gov)) for further information.  
 Trace considered overall poor quality  
 Insert length: 769 Std Error: 0.00  
 Seq primer: -40m13 fwd. Et from Amersham  
 High quality sequence stop: 1.

**FEATURES**  
 source

1. .37  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1503647"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NbHL19W, testis NHT, and B-cell  
 NCI\_CGAP\_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaído. "

BASE COUNT  
ORIGIN

```
Query Match      69.3%; Score 10.4; DB 13; Length 37;
Best Local Similarity 91.7%; Pred. No. 1.3e+05;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy	1	tggttatcttc	12
Db	34	TGTCCTATCTTC	23

```

RESULT 7
AA946896
LOCUS
DEFINITION
    AA946896 37 bp mRNA
    q53f01.s1 NCI.CCAP.Kid5 Homo sapiens cDNA clone IMAGE:159073 3'
    similar to TR:007708 O07708 HYPOTHETICAL 23.7 KD PROTEIN. ; mRNA
    sequence.
ACCESION
VERSION
KEYWORDS
SOURCE
    AA946896.1 GI:3110291
    EST.
    human.

```

**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 37)  
**AUTHORS** NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbbrp/image/image.html](http://www.bio.llnl.gov/dbbrp/image/image.html)  
Insert Length: 1106 Std Error: 0.00  
Seq primer: -40m13 fwd. E from Amersham  
High quality sequence stop: 1.

<b>FEATURES</b>	<b>SOURCE</b>
-----------------	---------------

```

location/qualifiers
1. .37
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1590073"
/clone_lib="NCI_CGAP_kid5"
/tissue_types="2_pooled tumors (clear cell type)"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer '[5',
AACTGGAGAATATTCGCGCGCGCAATATTTTTTTTTTTTTTTT 3']",
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonalido.
10 c 13 t
a 9 13 t

```

BASE COUNT  
ORIGIN

Query Match 69.38; Score 10.4; DB 14; Length 37;

```

Best Local Similarity 91.7%; Pred. No. 1.3e+05;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 cttatcttcagt 15
      |||||
Db 11 CTTATCTTCATT 22

```

RESULT	8
LOCUS	D45802/c
DEFINITION	D45802 38 bp mRNA EST 20-FEB-1995 HMG503019 Human adult lung 3' directed MboI cDNA Homo sapiens cDNA
ACCESSION	D45802
VERSION	D45802.1
KEYWORDS	GI:662756
SOURCE	EST. human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 38)  
Itoh, K., Okubo, K., Yosii, J., Yokouchi, H. and Matsubara, K.  
An expression profile of active genes in human lung  
DNA Research 1, 279-287 (1994)  
95236275  
Contact: Kohichi Itoh  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1, Yamadaoka, Suita, Osaka, 565, Japan  
Tel: 06-877-5111 x3910  
Fax: 06-877-1922.

<b>FEATURES</b>	<b>SOURCE</b>
-----------------	---------------

```

source
1..38
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human adult lung 3', directed MboI cDNA"
/notes="Adult human lung, 3', directed MboI"
12 a 4 c 5 g 17 t
BASE COUNT
ORIGIN

```

BASE COUN	ORIGIN
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

```
Query Match          69.3%; Score 10.4; DB 156; Length 38;
Best Local Similarity 91.7%; Pred. No. 1.3e+05;
Matches 11: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 3 gcttattctcag 14  
| | | | | | | |  
Db 20 GTTTATCTTCAG 9

RESULT 9  
AU010757

LOCUS	AU010757	42 bp	mRNA	EST	31-JUL-1998				
DEFINITION	AU010757 Schizosaccharomyces pombe late log phase cDNA								
	Schizosaccharomyces pombe cDNA clone spci0335, mRNA sequence.								
ACCESSION	AU010757								
VERSION	AU010757.1								
KEYWORDS	GI:3347437								
SOURCE	EST.								
	fission yeast.								
ORGANISM	Schizosaccharomyces pombe								

REFERENCE  
AUTHORS

**AUTHORS** Morimyo,M. and Mita,A.  
**TITLE** Identification of expressed sequence tags of Schizosaccharomyces pombe  
**JOURNAL** Unpublished (1998)  
**COMMENT** Contact: Mitsuoki Morimyo  
Genome Research Group  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-ku, Chiba 263-8555, Japan  
Email: morimyo@nirs.go.jp.

FEATURES  
source

Location/Qualifiers  
1. .42  
/organism="Schizosaccharomyces pombe"  
/strain="972"  
/db\_xref="taxon:4896"  
/clone="spc10335"  
/sex="h minus"  
/note="Vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, <http://www.nirs.go.jp>)"  
6 a 10 c 8 g 18 t

BASE COUNT  
ORIGIN

Query Match 69.3%; Score 10.4; DB 106; Length 42;  
Best Local Similarity 91.7%; Pred. No. 1.3e+05;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 gcttatcttcag 14  
||||| |||||  
Db 2 GCTATCTTCAG 13

RESULT 10  
AZ772066/c

LOCUS AZ772066 45 bp DNA GSS 16-FEB-2001  
DEFINITION IM0574L08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0574L08 R, DNA sequence.

ACCESSION AZ772066  
VERSION AZ772066.1 GI:12894990

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 45)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0574 row: L column: 08  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 45.  
Location/Qualifiers  
1. .45

FEATURES  
source

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0574L08"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114[gb]/AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 11 a 7 c 13 g 14 t  
ORIGIN

Query Match 69.3%; Score 10.4; DB 249; Length 45;  
Best Local Similarity 91.7%; Pred. No. 1.4e+05;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 gcttatcttcag 14  
||||| |||||  
Db 16 GCTTATGTTTCAG 5

RESULT 11  
AZ324419

LOCUS AZ324419 46 bp DNA GSS 29-SEP-2000  
DEFINITION IM0046B20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0046B20 F, DNA sequence.

ACCESSION AZ324419  
VERSION AZ324419.1 GI:10380118

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 46)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0046 row: B column: 20  
Seq primer: CGTGTAAACAGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 46.  
Location/Qualifiers  
1. .46

FEATURES  
source

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0046B20"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMDA2 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 14 a 14 c 6 g 12 t  
ORIGIN

Query Match 69.3%; Score 10.4; DB 241; Length 46;  
Best Local Similarity 91.7%; Pred. No. 1.4e+05; Length 46;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 cttatcttcag 15  
| | | | | | | | | |  
Db 25 CATATCTTCAGT 36

## RESULT 12

LOCUS A2793395 46 bp DNA 16-FEB-2001  
DEFINITION 2M0046D07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0046D07 R, DNA sequence.  
ACCESSION A2793395  
VERSION A2793395.1 GI:12938305  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus. 1 (bases 1 to 46)  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL plasmid inserts  
COMMENT Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0046 row: D column: 07  
Seq primer: CACACAGGAACACCTATGACC  
Class: plasmid ends  
High quality sequence stop: 46.

Location/Qualifiers  
1. 46  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0046D07"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMDA2nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson

FEATURES  
source

Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMDA2 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 15 a 8 c 3 g 20 t  
ORIGIN

Query Match 69.3%; Score 10.4; DB 249; Length 46;  
Best Local Similarity 91.7%; Pred. No. 1.4e+05; Length 46;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttc 12  
| | | | | | | | | |  
Db 19 TGGCTTACCTTC 30

## RESULT 13

LOCUS AW772664 49 bp mRNA EST 04-MAY-2000  
DEFINITION xn02b05.y1 NCI\_CGAP\_L15 Homo sapiens cDNA clone IMAGE:2692497 5', mRNA sequence.  
ACCESSION AW772664  
VERSION AW772664.1 GI:7704663  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 49)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: [info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: -40RP from Gibco.

Location/Qualifiers

1. 49  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2692497"  
/clone\_lib="NCI\_CGAP\_L15"  
/tissue\_type="hepatic adenoma"  
/lab\_host="DH10B"  
/note="Organ: liver; Vector: pCMV-SPORT4; Site:1: SalI;  
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 0.8 kb."  
22 a 14 c 6 g 7 t

BASE COUNT  
ORIGIN



Query Match 69.3%; Score 10.4; DB 120; Length 49;  
 Best Local Similarity 91.7%; Pred. No. 1.4e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggcttatcttca 13  
 | ||||| |||||  
 Db 9 GCCTTACTTCA 20

RESULT 14  
 AU103710/c  
 LOCUS  
 DEFINITION AU103710 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
 HEP14727, mRNA sequence.  
 ACCESSION AU103710  
 VERSION AU103710  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
 ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
 ,K., Suyama,A. and Sugano,S.  
 TITLE Fine structural analysis of transcription start sites of human  
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: yszuki@ims.u-tokyo.ac.jp  
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
 ,S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES  
 source  
 1..50  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HEP14727"  
 /clone\_lib="Sugano Homo sapiens cDNA library"  
 BASE COUNT 9 a 12 c 20 g 9 t  
 ORIGIN

Query Match 69.3%; Score 10.4; DB 107; Length 50;  
 Best Local Similarity 91.7%; Pred. No. 1.4e+05;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggcttatcttca 13  
 | ||||| |||||  
 Db 31 GGCTTACTTCA 20

RESULT 15  
 TA288E04P  
 LOCUS  
 DEFINITION TA288E04P 22 bp DNA GSS 13-DEC-2000  
 T. brucei sheared genomic DNA clone 288e04, forward sequence,  
 genomic survey sequence.  
 ACCESSION AL487174  
 VERSION AL487174.1 GI:11853828  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 REFERENCE 1 (bases 1 to 22)  
 AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,  
 Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,  
 Melville,S.E., Rajandream,M.A. and Barrell,B.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
 nh@sanger.ac.uk  
 Constructed at the Institute for Genomic Research (TIGR),  
 Rockville, MD. Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
 to give a tight size distribution (4 kb). The v + i method used for the library construction is  
 described in detail in Smith, H. and Venter, J.C. (Making small  
 insert libraries for whole genome shotgun sequencing projects. In  
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Barrell, Oxford University Press, 1999).  
 Email: nelsayed@tigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available  
 at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

FEATURES  
 source  
 1..22

/organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="288e04"  
 BASE COUNT 4 a 6 c 4 g 8 t  
 ORIGIN

Query Match 68.0%; Score 10.2; DB 258; Length 22;  
 Best Local Similarity 80.0%; Pred. No. 1.6e+05;  
 Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15  
 | ||||| |||||  
 Db 7 TCGCTTATGTTCCGT 21

Search completed: October 2, 2001, 15:01:15  
 Job time: 10844 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:55:01 ; Search time 2173.58 Seconds  
(without alignments)  
19.290 Million cell updates/sec

Title: US-09-757-100B-10

Perfect score: 20

Sequence: 1 ggtgaagcagctgcattatt 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1895189 seqs, 1048201267 residues

Total number of hits satisfying chosen parameters: 299022

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_New:\*

- 1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*
- 8: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	12.8	64.0	17	6	US-09-866-108-7017
C 2	12.8	64.0	17	6	US-09-866-108-7018
C 3	12.8	64.0	25	6	US-09-866-108-11909
C 4	12.8	64.0	25	6	US-09-866-108-11910
C 5	12.8	64.0	25	6	US-09-866-108-11911
C 6	12.8	64.0	25	6	US-09-866-108-11912
C 7	12.8	64.0	25	6	US-09-866-108-11913
C 8	12.8	64.0	25	6	US-09-866-108-11914
C 9	12.8	64.0	25	6	US-09-866-108-11915
C 10	12.8	64.0	25	6	US-09-866-108-11916
C 11	12.8	64.0	25	6	US-09-866-108-11917
C 12	12.8	64.0	25	6	US-09-866-108-11918
C 13	12.2	61.0	33	5	US-09-930-251-10
C 14	12.2	61.0	42	8	US-60-278-561-6257
C 15	12.2	61.0	49	7	US-09-681-508-1
C 16	12	60.0	17	6	US-09-866-108-7019
C 17	12	60.0	17	6	US-09-866-108-7020
C 18	12	60.0	17	6	US-09-866-108-7021
C 19	12	60.0	17	6	US-09-866-108-7022
C 20	12	60.0	25	6	US-09-866-108-11919
C 21	12	60.0	25	6	US-09-866-108-11920
C 22	12	60.0	25	6	US-09-866-108-11921
C 23	12	60.0	25	6	US-09-866-108-11922
C 24	12	60.0	44	8	US-60-253-456-27612
C 25	11.8	59.0	17	6	US-09-866-108-7016

ALIGNMENTS

RESULT 1

US-09-866-108-7017/c	17	7	US-09-818-875-2710	Sequence 2710, Ap
Sequence 7017, Application US/09866108	17	7	US-09-818-875-2711	Sequence 2711, Ap
GENERAL INFORMATION:	17	7	US-09-818-875-2714	Sequence 2714, Ap
APPLICANT: GU, Yizhong	17	7	US-09-818-875-2715	Sequence 2715, Ap
APPLICANT: JI, Yonggang	25	6	US-09-866-108-11908	Sequence 11908, A
APPLICANT: PENN, Shanton G.	27	7	US-09-772-315-2	Sequence 2, Appl
APPLICANT: HANZEL, David K.	38	5	US-09-371-772B-7807	Sequence 7807, Ap
APPLICANT: RANK, David R.	38	5	US-09-708-690-10084	Sequence 10084, A
APPLICANT: CHEN, Wensheng	43	8	US-60-253-654-13218	Sequence 13218, A
APPLICANT: SHANNON, Mark	43	8	US-60-255-592-13218	Sequence 13218, A
TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE	29	7	US-09-214-916A-7	Sequence 7, Appl
FILE REFERENCE: AEOMICA-7	40	6	US-09-803-736-1511	Sequence 1511, Ap
CURRENT APPLICATION NUMBER: US/09/866,108	50	6	US-09-504-576A-14337	Sequence 14337, A
CURRENT FILING DATE: 2001-05-25	20	5	US-09-923-515-13	Sequence 13, Appl
PRIOR APPLICATION NUMBER: US 60/207,456	20	5	US-09-923-515-15	Sequence 15, Appl
PRIOR FILING DATE: 2000-05-26	20	5	US-09-923-515-18	Sequence 18, Appl
PRIOR APPLICATION NUMBER: GB 24263.6	20	6	US-09-198-452A-2269	Sequence 2269, Ap
PRIOR FILING DATE: 2000-10-04	20	6	US-09-198-452A-4764	Sequence 4764, Ap
PRIOR APPLICATION NUMBER: US 60/236,359	20	7	US-09-659-845A-98	Sequence 98, Appl
PRIOR FILING DATE: 2000-09-27	47	5	US-09-877-804-10	Sequence 10, Appl
PRIOR APPLICATION NUMBER: PCT/US01/00666				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00667				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00664				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00669				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00665				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00668				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00663				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00662				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00661				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00670				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: US 60/234,687				
PRIOR FILING DATE: 2000-09-21				
PRIOR APPLICATION NUMBER: US 60/266,860				
PRIOR FILING DATE: 2001-02-05				
NUMBER OF SEQ ID NOS: 15752				
SOFTWARE: Aeomica Sequence Listing Engine				
SEQ ID NO 7017				

;  
; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-7017

Query Match 64.0%; Score 12.8; DB 6; Length 17;  
Best Local Similarity 87.5%; Pred. No. 3e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 aagcagctgccattat 19  
|||||  
Db 17 AAGCAGCTGCCACCAT 2

## RESULT 2

US-09-866-108-7018/c  
; Sequence 7018, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.  
; APPLICANT: HANZEL, David K.  
; APPLICANT: RANK, David R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: SHANNON, Mark  
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
; FILE REFERENCE: AEOMICA-7  
; CURRENT APPLICATION NUMBER: US/09/866,108  
; CURRENT FILING DATE: 2001-05-25  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 60/266,860  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aeomica Sequence Listing Engine  
; SEQ ID NO 7018  
; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-7018

Query Match 64.0%; Score 12.8; DB 6; Length 17;  
Best Local Similarity 87.5%; Pred. No. 3e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 aagcagctgccattat 19  
|||||  
Db 16 AAGCAGCTGCCACCAT 1

## RESULT 3

US-09-866-108-11909/c  
; Sequence 11909, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.  
; APPLICANT: HANZEL, David K.  
; APPLICANT: RANK, David R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: SHANNON, Mark  
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
; FILE REFERENCE: AEOMICA-7  
; CURRENT APPLICATION NUMBER: US/09/866,108  
; CURRENT FILING DATE: 2001-05-25  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 60/266,860  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aeomica Sequence Listing Engine  
; SEQ ID NO 11909  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-11909

Query Match 64.0%; Score 12.8; DB 6; Length 25;  
Best Local Similarity 87.5%; Pred. No. 3.2e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 aagcagctgccattat 19  
|||||  
Db 25 AAGCAGCTGCCACCAT 10

## RESULT 4

US-09-866-108-11910/c  
; Sequence 11910, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong

```

; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aecomica Sequence Listing Engine.
; SEQ ID NO 11910
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-11910

Query Match 64.0%; Score 12.8; DB 6; Length 25;
Best Local Similarity 87.5%; Pred. No. 3.2e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps

QY 4 aagcagctgccattat 19
Db 24 AAGCAGCTGCCACCAT 9
|||||
|||||

RESULT 5
US-09-866-108-11911/c
; Sequence 11911, Application US/09866108
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25

```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aecomica Sequence Listing Engine
; SEQ ID NO 11912
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-11912
```

```
Query Match 64.0%; Score 12.8; DB 6; Length 25;
Best Local Similarity 87.5%; Pred. No. 3.2e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 4 aagcagctgccattat 19
||||| ||||| ||
Db 22 AAGCAGCTGCCACCAT 7
```

```
RESULT 7
US-09-866-108-11913/c
; Sequence 11913, Application US/09866108
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AECOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aecomica Sequence Listing Engine
; SEQ ID NO 11913
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-11913
```

```
Query Match 64.0%; Score 12.8; DB 6; Length 25;
Best Local Similarity 87.5%; Pred. No. 3.2e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 4 aagcagctgccattat 19
||||| ||||| ||
Db 21 AAGCAGCTGCCACCAT 6
```

```
RESULT 8
US-09-866-108-11914/c
; Sequence 11914, Application US/09866108
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AECOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
```

; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aecomica Sequence Listing Engine  
; SEQ ID NO 11914  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-11914

Query Match 64.0%; Score 12.8; DB 6; Length 25;  
Best Local Similarity 87.5%; Pred. No. 3.2e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 aagcagctgccattat 19  
|||||  
Db 20 AAGCAGCTGCCACCAT 5

## RESULT 9

US-09-866-108-11915/c  
; Sequence 11915, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.  
; APPLICANT: HANZEL, David K.  
; APPLICANT: RANK, David R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: SHANNON, Mark  
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
; FILE REFERENCE: AECOMICA-7  
; CURRENT APPLICATION NUMBER: US/09/866,108  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00671  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00672  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 60/266,860  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aecomica Sequence Listing Engine  
; SEQ ID NO 11915  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-11915

Query Match 64.0%; Score 12.8; DB 6; Length 25;

Best Local Similarity 87.5%; Pred. No. 3.2e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 4 aagcagctgccattat 19  
|||||  
Db 19 AAGCAGCTGCCACCAT 4

## RESULT 10

US-09-866-108-11916/c  
; Sequence 11916, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.  
; APPLICANT: HANZEL, David K.  
; APPLICANT: RANK, David R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: SHANNON, Mark  
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
; FILE REFERENCE: AECOMICA-7  
; CURRENT APPLICATION NUMBER: US/09/866,108  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 60/266,860  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aecomica Sequence Listing Engine  
; SEQ ID NO 11916  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-11916

Query Match 64.0%; Score 12.8; DB 6; Length 25;  
Best Local Similarity 87.5%; Pred. No. 3.2e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 aagcagctgccattat 19  
|||||  
Db 18 AAGCAGCTGCCACCAT 3

RESULT 11  
US-09-866-108-11917/c

```

US-09-866-108-11918/c
: Sequence 11918, Application US/09866108
:
: GENERAL INFORMATION:
:
: APPLICANT: GU, Yizhong
: APPLICANT: JI, Yonggang
: APPLICANT: PENN, Sharron G.
: APPLICANT: HANZEL, David K.
: APPLICANT: RANK, David R.
: APPLICANT: CHEN, Wensheng
: APPLICANT: SHANNON, Mark
: TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
:

```

Query Match 61.0%; Score 12.2; DB 5; Length 33;  
Best Local Similarity 82.4%; Pred. No. 7.1e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels



QY 4 aagcagctgccattatt 20  
|| ||| ||||| |||||  
Db 16 aaacagatgccagtatt 32

## RESULT 14

US-60-278-561-6257/c  
; Sequence 6257, Application US/60278561  
; GENERAL INFORMATION:  
; APPLICANT: Morris, MacDonald  
; APPLICANT: Lal, Preeti  
; APPLICANT: Diep, Dinh  
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using  
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms Identified Thereby  
; FILE REFERENCE: GX-0012-1 P  
; CURRENT APPLICATION NUMBER: US/60/278,561  
; CURRENT FILING DATE: 2001-03-23  
; NUMBER OF SEQ ID NOS: 15598  
; SOFTWARE: PERL Program  
; SEQ ID NO 6257  
; LENGTH: 42  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No: SNP00060057  
; NAME/KEY: snp  
; LOCATION: 26  
; OTHER INFORMATION: 229652.27, 1901, A->G  
US-60-278-561-6257

Query Match 61.0%; Score 12.2; DB 8; Length 42;  
Best Local Similarity 82.4%; Pred. No. 7.5e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 taagcagctgccattat 19  
| ||||| |||||  
Db 30 TCAGCAGCTGGGATTAT 14

## RESULT 15

US-09-681-508-1/c  
; Sequence 1, Application US/09681508  
; GENERAL INFORMATION:  
; APPLICANT: Lin, Yun  
; APPLICANT: Heil, James R  
; APPLICANT: Jayasena, Sumedha D  
; TITLE OF INVENTION: Aptamer Based Two-Site Binding Assay  
; FILE REFERENCE: NEX 89  
; CURRENT APPLICATION NUMBER: US/09/681,508  
; CURRENT FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,016  
; PRIOR FILING DATE: 2000-04-18  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 49  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Nucleic Acid Ligand  
US-09-681-508-1

Query Match 61.0%; Score 12.2; DB 7; Length 49;  
Best Local Similarity 82.4%; Pred. No. 7.7e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gtaagcagctgccatta 18

Db 49 GTAAGCGAAGCCATTA 33  
||||| | |||||

Search completed: October 2, 2001, 16:55:02  
Job time: 17665 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 21:50:11 ; Search time 17695.9 Seconds  
(without alignments)  
16.701 Million cell updates/sec

Title: US-09-757-100B-9

Perfect score: 20

Sequence: 1 ttgtgtagtagtagtagtat 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1415048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 4519004

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:\*

- 1: /cgn2\_6/ptodata/1/pna/US0612\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pna/US0613\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq:\*
- 20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq:\*
- 21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq:\*
- 22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq:\*
- 23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq:\*
- 24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq:\*
- 25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq:\*
- 26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq:\*
- 27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq:\*
- 28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq:\*
- 29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq:\*
- 30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq:\*
- 31: /cgn2\_6/ptodata/1/pna/US098\_COMB.seq:\*
- 32: /cgn2\_6/ptodata/1/pna/US099\_COMB.seq:\*
- 33: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq:\*
- 34: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq:\*
- 35: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq:\*
- 36: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq:\*
- 37: /cgn2\_6/ptodata/1/pna/US099E\_COMB.seq:\*
- 38: /cgn2\_6/ptodata/1/pna/US099F\_COMB.seq:\*
- 39: /cgn2\_6/ptodata/1/pna/US099G\_COMB.seq:\*
- 40: /cgn2\_6/ptodata/1/pna/US099H\_COMB.seq:\*
- 41: /cgn2\_6/ptodata/1/pna/US099I\_COMB.seq:\*
- 42: /cgn2\_6/ptodata/1/pna/US099J\_COMB.seq:\*
- 43: /cgn2\_6/ptodata/1/pna/US099K\_COMB.seq:\*

- 44: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq:\*
- 45: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq:\*
- 46: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq:\*
- 47: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq:\*
- 48: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq:\*
- 49: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq:\*
- 50: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq:\*
- 51: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq:\*
- 52: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq:\*
- 53: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq:\*
- 54: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq:\*
- 55: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq:\*
- 56: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq:\*
- 57: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq:\*
- 58: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq:\*
- 59: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq:\*
- 60: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	20	1	PCT-US00-18999-9
2	20	100.0	20	17	US-09-377-310-9
3	20	100.0	20	29	US-09-757-100B-9
4	15	75.0	15	1	PCT-US00-18999-29
5	15	75.0	15	17	US-09-377-310-29
6	15	75.0	15	29	US-09-757-100B-29
7	15	75.0	20	1	PCT-US00-18999-8
8	15	75.0	20	17	US-09-377-310-8
9	15	75.0	20	29	US-09-757-100B-8
10	14.8	74.0	47	18	US-09-422-978-407
11	14.4	72.0	25	55	US-60-232-638-58459
12	14.4	72.0	25	55	US-60-232-638-16838
13	14.2	71.0	25	55	US-60-234-017-209206
14	13.8	69.0	20	18	US-09-486-816-26
15	13.8	69.0	25	55	US-60-232-638-50420
16	13.8	69.0	36	31	US-09-449-631-83
17	13.8	69.0	36	31	US-09-848-616-83
18	13.6	68.0	25	55	US-60-232-638-45369
19	13.6	68.0	50	14	US-09-040-266-1519
20	13.4	67.0	47	18	US-09-422-978-537
21	13.2	66.0	25	55	US-60-233-166-183276
22	13.2	66.0	25	55	US-60-233-166-365262
23	13.2	66.0	25	55	US-60-234-017-181973
24	13.2	66.0	25	55	US-60-234-017-181978
25	13.2	66.0	31	16	US-09-238-402-550
26	13	65.0	15	1	PCT-US00-18999-28
27	13	65.0	15	17	US-09-377-310-28
28	13	65.0	15	29	US-09-757-100B-28
29	13	65.0	25	55	US-60-232-638-44200
30	13	65.0	25	55	US-60-232-638-4223
31	12.8	64.0	25	55	US-60-232-638-121622
32	12.8	64.0	25	55	US-60-233-166-359242
33	12.8	64.0	25	55	US-60-233-166-365273
34	12.8	64.0	25	55	US-60-233-166-9949
35	12.8	64.0	25	55	US-60-233-620-84153
36	12.8	64.0	25	55	US-60-234-049-44209
37	12.6	63.0	25	26	US-09-660-220-36361
38	12.6	63.0	25	55	US-60-232-638-21996
39	12.6	63.0	25	55	US-60-232-638-45114
40	12.6	63.0	25	55	US-60-232-638-130842
41	12.6	63.0	25	55	US-60-233-620-27969
42	12.6	63.0	25	55	US-60-234-017-352215
43	12.6	63.0	25	55	US-60-234-049-85629
44	12.6	63.0	36	1	PCT-US99-18603-15
45	12.6	63.0	36	1	PCT-US99-18603-16

## ALIGNMENTS

RESULT 1  
 PCT-US00-18999-9  
 ; Sequence 9, Application PC/TUS0018999  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Isis Pharmaceuticals, Inc.  
 ; APPLICANT: Monia, Brett P.  
 ; APPLICANT: Gaarde, William A.  
 ; APPLICANT: Nero, Pamela S.  
 ; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
 ; FILE REFERENCE: ISPH-0476  
 ; CURRENT APPLICATION NUMBER: PCT/US00/18999  
 ; PRIOR FILING DATE: 2000-07-13  
 ; PRIOR APPLICATION NUMBER: 09/377,310  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 20  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: antisense sequence  
 PCT-US00-18999-9

Query Match 100.0%; Score 20; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttgcctagctaggtat 20  
 |||||  
 Db 1 ttttgcctagctaggtat 20

RESULT 2  
 US-09-377-310-9  
 ; Sequence 9, Application US/09377310A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Monia, Brett P.  
 ; APPLICANT: Gaarde, William A.  
 ; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
 ; FILE REFERENCE: ISPH-0389  
 ; CURRENT APPLICATION NUMBER: US/09/377,310A  
 ; PRIOR FILING DATE: 1999-08-19  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 20  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: antisense sequence  
 US-09-377-310-9

Query Match 100.0%; Score 20; DB 17; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttgcctagctaggtat 20  
 |||||  
 Db 1 ttttgcctagctaggtat 20

RESULT 3  
 US-09-757-100b-9  
 ; Sequence 9, Application US/09757100B

; GENERAL INFORMATION:  
 ; APPLICANT: Monia, Brett P.  
 ; APPLICANT: Gaarde, William A.  
 ; APPLICANT: Nero, Pamela S.  
 ; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
 ; FILE REFERENCE: ISPH-0533  
 ; CURRENT APPLICATION NUMBER: US/09/757,100B  
 ; PRIOR FILING DATE: 2001-03-15  
 ; PRIOR APPLICATION NUMBER: 09/377,310  
 ; PRIOR FILING DATE: 1999-08-19  
 ; PRIOR APPLICATION NUMBER: PCT/US00/18999  
 ; PRIOR FILING DATE: 2000-07-13  
 ; NUMBER OF SEQ ID NOS: 44  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 20  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: antisense sequence  
 US-09-757-100B-9

Query Match 100.0%; Score 20; DB 29; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttgcctagctaggtat 20  
 |||||  
 Db 1 ttttgcctagctaggtat 20

RESULT 4  
 PCT-US00-18999-29  
 ; Sequence 29, Application PC/TUS0018999  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Isis Pharmaceuticals, Inc.  
 ; APPLICANT: Monia, Brett P.  
 ; APPLICANT: Gaarde, William A.  
 ; APPLICANT: Nero, Pamela S.  
 ; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
 ; FILE REFERENCE: ISPH-0476  
 ; CURRENT APPLICATION NUMBER: PCT/US00/18999  
 ; PRIOR FILING DATE: 2000-07-13  
 ; PRIOR APPLICATION NUMBER: 09/377,310  
 ; PRIOR FILING DATE: 1999-08-19  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 29  
 ; LENGTH: 15  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: antisense sequence  
 PCT-US00-18999-29

Query Match 75.0%; Score 15; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ttgcctagctaggtat 17  
 |||||  
 Db 1 ttgcctagctaggtat 15

RESULT 5  
 US-09-377-310-29  
 ; Sequence 29, Application US/09377310A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Monia, Brett P.

; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-29

Query Match 75.0%; Score 15; DB 17; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 ttgctagatgctagg 17  
Db 1 ttgctagatgctagg 15

RESULT 6  
US-09-757-100B-29  
; Sequence 29, Application US/09757100B  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100B-29

Query Match 75.0%; Score 15; DB 29; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 ttgctagatgctagg 17  
Db 1 ttgctagatgctagg 15

RESULT 7  
PCT-US00-18999-8  
; Sequence 8, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0476

; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; CURRENT FILING DATE: 2000-07-13  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-8

Query Match 75.0%; Score 15; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 6 ctgatgctaggtat 20  
Db 1 ctgatgctaggtat 15

RESULT 8  
US-09-377-310-8  
; Sequence 8, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-8

Query Match 75.0%; Score 15; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 6 ctgatgctaggtat 20  
Db 1 ctgatgctaggtat 15

RESULT 9  
US-09-757-100B-8  
; Sequence 8, Application US/09757100B  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 8
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: antisense sequence
US-09-757-100B-8

Query Match          75.0%; Score 15; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ctatagctaggtat 20
   ||||| ||||| |||||
Db 1 ctatagctaggtat 15

RESULT 10
US-09-422-978-407
; Sequence 407, Application US/09422978
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
; FILE REFERENCE: GENSET.020CPI
; CURRENT APPLICATION NUMBER: US/09/422,978
; CURRENT FILING DATE: 1999-10-20
; EARLIER APPLICATION NUMBER: US 09/298,850
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,732
; EARLIER FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: US 60/082,614
; EARLIER FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 407
; LENGTH: 47
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 24
; OTHER INFORMATION: 99-14678-75 : polymorphic base G or C
US-09-422-978-407

Query Match          74.0%; Score 14.8; DB 18; Length 47;
Best Local Similarity 88.9%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ttgctagatgctaggtat 20
   |||| | ||||| ||||| ||
Db 4 ttgcaagatcgtaggaaat 21

RESULT 11
US-60-232-638-58459/c
; Sequence 58459, Application US/60232638
; GENERAL INFORMATION:
; APPLICANT: Mittmann
; APPLICANT: Affymetrix, Inc.
; TITLE OF INVENTION: Methods of Genetic Analysis of Yeast
; FILE REFERENCE: 3110
; CURRENT APPLICATION NUMBER: US/60/232,638
; CURRENT FILING DATE: 2000-09-14
; NUMBER OF SEQ ID NOS: 138410
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 58459
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; PUBLICATION INFORMATION:
```

```
; DATABASE ACCESSION NUMBER: SGD YER050C
US-60-232-638-58459

Query Match          72.0%; Score 14.4; DB 55; Length 25;
Best Local Similarity 93.8%; Pred. No. 3.4e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tgctagatgctaggtat 19
   || ||||| ||||| |||||
Db 24 TGATAGATGCTAGGTA 9

RESULT 12
US-60-233-620-16838
; Sequence 16838, Application US/60233620
; GENERAL INFORMATION:
; APPLICANT: Mittmann
; APPLICANT: Affymetrix, Inc.
; TITLE OF INVENTION: Methods of Genetic Analysis of
; TITLE OF INVENTION: Arabidopsis thaliana
; FILE REFERENCE: 3116
; CURRENT APPLICATION NUMBER: US/60/233,620
; CURRENT FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 131820
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16838
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank AC005499
US-60-233-620-16838

Query Match          72.0%; Score 14.4; DB 55; Length 25;
Best Local Similarity 93.8%; Pred. No. 3.4e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tgctagatgctaggtat 19
   ||||| ||||| ||||| ||
Db 1 tgctagaagctaggtat 16

RESULT 13
US-60-234-017-209206/c
; Sequence 209206, Application US/60234017
; GENERAL INFORMATION:
; APPLICANT: Mittmann, M
; APPLICANT: Affymetrix, Inc.
; TITLE OF INVENTION: Methods of Genetic Analysis of Mus
; TITLE OF INVENTION: musculus
; FILE REFERENCE: 3115
; CURRENT APPLICATION NUMBER: US/60/234,017
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 605887
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 209206
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank AI504141
US-60-234-017-209206

Query Match          71.0%; Score 14.2; DB 55; Length 25;
Best Local Similarity 84.2%; Pred. No. 4.3e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtctagatgctaggtat 20
   |||| | ||||| |||||
Db 23 TTTGTTAGATGCTCGGTAT 5
```

Search completed: October 2, 2001, 21:50:12  
Job time: 24525 sec

RESULT 14  
US-09-486-816-26  
; Sequence 26, Application US/09486816  
; GENERAL INFORMATION:  
; APPLICANT: HARADA, Hideyuki  
; APPLICANT: FUJIWARA, Masanori  
; APPLICANT: TAGASHIRA, Shuzo  
; APPLICANT: OGAWA, Shinji  
; APPLICANT: KATSUMATA, Takashi  
; APPLICANT: NAKATSUKA, Masashi  
; TITLE OF INVENTION: DNA Participating In Regulation of Expression of PEBP2alpha Ge  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: 0020-4685P  
; CURRENT APPLICATION NUMBER: US/09/486,816  
; CURRENT FILING DATE: 2000-05-30  
; PRIOR APPLICATION NUMBER: 254250/1997 Japan  
; PRIOR FILING DATE: 1997-09-02  
; PRIOR APPLICATION NUMBER: 299407/1997 Japan  
; PRIOR FILING DATE: 1997-10-15  
; PRIOR APPLICATION NUMBER: 114135/1998 Japan  
; PRIOR FILING DATE: 1998-04-08  
; PRIOR APPLICATION NUMBER: PCT/JP98/03920  
; PRIOR FILING DATE: 1998-09-02  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 26  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Other nucleic acid, synthetic DNA  
US-09-486-816-26

Query Match 69.0%; Score 13.8; DB 18; Length 20;  
Best Local Similarity 88.2%; Pred. No. 6.6e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 2 ttgtctagatgctaggt 18  
||| ||| ||| ||| ||| |||  
Db 3 ttgtataaatgctaggt 19

RESULT 15  
US-60-232-638-50420/C  
; Sequence 50420, Application US/60232638  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of Yeast  
; FILE REFERENCE: 3110  
; CURRENT APPLICATION NUMBER: US/60/232,638  
; CURRENT FILING DATE: 2000-09-14  
; NUMBER OF SEQ ID NOS: 138410  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 50420  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: SGD YDR240C  
US-60-232-638-50420

Query Match 69.0%; Score 13.8; DB 55; Length 25;  
Best Local Similarity 88.2%; Pred. No. 6.8e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 2 ttgtctagatgctaggt 18  
||| ||| ||| ||| ||| |||  
Db 22 TTACTAGATGCTAGTT 6





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:55:01 ; Search time 2173.58 Seconds  
(without alignments)  
19.290 Million cell updates/sec

Title: US-09-757-100B-9

Perfect score: 20

Sequence: 1 ttttctagatgctaggtat 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1895189 seqs, 1048201267 residues

Total number of hits satisfying chosen parameters: 299022

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_New.\*

- 1: /cgn2.6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*
- 2: /cgn2.6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*
- 3: /cgn2.6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*
- 4: /cgn2.6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*
- 5: /cgn2.6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 6: /cgn2.6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 7: /cgn2.6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 8: /cgn2.6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	13.8	69.0	45	US-60-278-258-6107	Sequence 6107, Ap
C 2	13.2	66.0	31	US-09-574-376B-1127	Sequence 1127, Ap
C 3	12.8	64.0	31	US-09-801-274-1141	Sequence 1141, Ap
C 4	12.8	64.0	31	US-09-801-274-1801	Sequence 1801, Ap
C 5	12.8	64.0	31	US-09-817-879-5496	Sequence 5496, Ap
C 6	12.4	62.0	27	US-09-445-283B-47	Sequence 47, Appl
C 7	12.2	61.0	31	US-09-817-879-4811	Sequence 4811, Ap
C 8	12.2	61.0	31	US-09-817-879-5557	Sequence 5557, Ap
C 9	12.2	61.0	31	US-09-817-879-6920	Sequence 6920, Ap
C 10	12.2	61.0	31	US-09-817-879-9054	Sequence 9054, Ap
C 11	12	60.0	36	US-60-253-378-11076	Sequence 11076, Ap
C 12	11.8	59.0	31	US-09-708-690-16613	Sequence 16613, A
C 13	11.8	59.0	36	US-09-232-170-20	Sequence 20, Appl
C 14	11.8	59.0	38	US-09-840-424-773	Sequence 773, Appl
C 15	11.6	58.0	31	US-09-801-274-1751	Sequence 1751, Ap
C 16	11.6	58.0	49	US-60-253-654-12658	Sequence 12658, A
C 17	11.6	58.0	49	US-60-255-592-12658	Sequence 12658, A
C 18	11.4	57.0	15	US-09-895-007-96	Sequence 96, Appl
C 19	11.4	57.0	17	US-09-920-313-96	Sequence 96, Appl
C 20	11.4	57.0	17	US-09-817-879-2124	Sequence 2124, Ap
C 21	11.4	57.0	17	US-09-817-879-2431	Sequence 2431, Ap
C 22	11.4	57.0	20	US-09-827-383-1410	Sequence 1410, Ap
C 23	11.4	57.0	33	US-09-555-998B-15	Sequence 15, Appl
C 24	11.4	57.0	40	US-60-278-232-4276	Sequence 4276, Ap
C 25	11.4	57.0	40	US-60-253-457-8122	Sequence 8122, Ap

ALIGNMENTS

RESULT 1

US-60-278-258-6107/c

Sequence 6107, Application US/60278258

GENERAL INFORMATION:

APPLICANT: Morris, MacDonald

APPLICANT: Lal, Preeti

APPLICANT: Diep, Dinh

TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using Polynucleotide Sequence Databases, and Single Nucleotide Polymorphisms Identified Thereby

FILE REFERENCE: GX-0010-1 P

CURRENT APPLICATION NUMBER: US/60/278,258

CURRENT FILING DATE: 2001-03-23

NUMBER OF SEQ ID NOS: 17730

SOFTWARE: PERL Program

SEQ ID NO 6107

LENGTH: 45

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc\_feature

OTHER INFORMATION: Incyte ID No: SNP00057112

NAME/KEY: snp

LOCATION: 20

OTHER INFORMATION: 350403.1, 20, G->A

US-60-278-258-6107

Query Match 69.0%; Score 13.8; DB 8; Length 45;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 tgcctagatgctaggtat 20

|||||

Db 20 TGCTAGATGGTAGTAT 4

RESULT 2

US-09-574-376B-1127/c

Sequence 1127, Application US/09574376B

GENERAL INFORMATION:

APPLICANT: Warrington, Janet

APPLICANT: Shah, Nila

APPLICANT: Gingers, Thomas Raymond

APPLICANT: Affymetrix, Inc.

TITLE OF INVENTION: Human Lymphoblast Polymorphisms

FILE REFERENCE: 3229.2

CURRENT APPLICATION NUMBER: US/09/574,376B

Sequence 2052, Ap  
Sequence 16945, A  
Sequence 17052, A  
Sequence 17300, A  
Sequence 18915, A  
Sequence 20162, A  
Sequence 5680, Ap  
Sequence 6194, Ap  
Sequence 6484, Ap  
Sequence 6683, Ap  
Sequence 7183, Ap  
Sequence 7502, Ap  
Sequence 8470, Ap  
Sequence 25152, A  
Sequence 191, App  
Sequence 379, App  
Sequence 24, Appl  
Sequence 6, Appl  
Sequence 3901, Ap  
Sequence 13773, A

; CURRENT FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 1330  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1127  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; PUBLIC INFORMATION:  
US-09-574-376B-1127

Query Match 66.0%; Score 13.2; DB 5; Length 31;  
Best Local Similarity 75.0%; Pred. No. 1.7e+03;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttttgcctagctaggtat 20  
|||||:|||||:|  
Db 24 TTCTGCTCRATGCTAGAGAT 5

RESULT 3  
US-09-801-274-1141  
; Sequence 1141, Application US/09801274  
; GENERAL INFORMATION:  
; APPLICANT: Cargill, Michele  
; APPLICANT: Ireland, James S.  
; APPLICANT: Lander, Eric S.  
; TITLE OF INVENTION: HUMAN SINGLE NUCLEOTIDE POLYMORPHISMS  
; FILE REFERENCE: 2825.2009-001  
; CURRENT APPLICATION NUMBER: US/09/801,274  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 60/187,510  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: US 60/206,129  
; PRIOR FILING DATE: 2000-05-22  
; NUMBER OF SEQ ID NOS: 1802  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1141  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-801-274-1141

Query Match 64.0%; Score 12.8; DB 6; Length 31;  
Best Local Similarity 77.8%; Pred. No. 2.8e+03;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttgctagatgctaggtat 20  
|||||:|||||  
Db 8 ttgccaaaygctcggtat 25

RESULT 4  
US-09-801-274-1801  
; Sequence 1801, Application US/09801274  
; GENERAL INFORMATION:  
; APPLICANT: Cargill, Michele  
; APPLICANT: Ireland, James S.  
; APPLICANT: Lander, Eric S.  
; TITLE OF INVENTION: HUMAN SINGLE NUCLEOTIDE POLYMORPHISMS  
; FILE REFERENCE: 2825.2009-001  
; CURRENT APPLICATION NUMBER: US/09/801,274  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 60/187,510  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: US 60/206,129  
; PRIOR FILING DATE: 2000-05-22  
; NUMBER OF SEQ ID NOS: 1802  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1801  
; LENGTH: 31  
; TYPE: DNA

; ORGANISM: Homo sapiens  
US-09-801-274-1801

Query Match 64.0%; Score 12.8; DB 6; Length 31;  
Best Local Similarity 77.8%; Pred. No. 2.8e+03;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttgctagatgctaggtat 20  
|||||:|||||  
Db 8 ttgccaaaygctcggtat 25

RESULT 5  
US-09-817-879-5496  
; Sequence 5496, Application US/09817879  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals Inc.  
; TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Rel  
; TITLE OF INVENTION: Hepatitis C Virus Infection  
; FILE REFERENCE: MBH00-801-F  
; CURRENT APPLICATION NUMBER: US/09/817,879  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 9703  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5496  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: DNazyme  
US-09-817-879-5496

Query Match 64.0%; Score 12.8; DB 6; Length 31;  
Best Local Similarity 87.5%; Pred. No. 2.8e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 tgctagatgctaggtat 19  
|||||:|||||  
Db 2 tgctagaggctagctat 17

RESULT 6  
US-09-445-283B-47/c  
; Sequence 47, Application US/09445283B  
; GENERAL INFORMATION:  
; APPLICANT: Malliga, Pal  
; APPLICANT: Silhavy, Daniel  
; APPLICANT: Sriraman, Priya  
; TITLE OF INVENTION: Plastid promoters for Transgene  
; TITLE OF INVENTION: Expression in the Plastids of Higher Plants  
; FILE REFERENCE: Rut 97-0097  
; CURRENT APPLICATION NUMBER: US/09/445,283B  
; CURRENT FILING DATE: 1999-12-03  
; PRIOR APPLICATION NUMBER: PCT/US98/11437  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/058,670  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/048,376  
; PRIOR FILING DATE: 1997-06-03  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 47  
; LENGTH: 27  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Primer  
US-09-445-283B-47

Query Match 62.0%; Score 12.4; DB 6; Length 27;  
Best Local Similarity 92.9%; Pred. No. 4.5e+03;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttgctagatgctag 16  
| | | | | | | | | |  
Db 17 TTGCTAGATTCTAG 4

## RESULT 7

US-09-817-879-4811  
; Sequence 4811, Application US/09817879  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals Inc.  
; TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Related  
; FILE REFERENCE: MH000-801-F  
; CURRENT APPLICATION NUMBER: US/09/817,879  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 9703  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4811  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: DNazyme  
US-09-817-879-4811

Query Match 61.0%; Score 12.2; DB 6; Length 31;  
Best Local Similarity 82.4%; Pred. No. 5.8e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttgctagatgctagta 19  
| | | | | | | | | |  
Db 1 ttgctagaggtagta 17

## RESULT 8

US-09-817-879-5557  
; Sequence 5557, Application US/09817879  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals Inc.  
; TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Related  
; FILE REFERENCE: MH000-801-F  
; CURRENT APPLICATION NUMBER: US/09/817,879  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 9703  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5557  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: DNazyme  
US-09-817-879-5557

Query Match 61.0%; Score 12.2; DB 6; Length 31;  
Best Local Similarity 82.4%; Pred. No. 5.8e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttgctagatgctagta 19  
| | | | | | | | | |  
Db 1 ttgctagaggtagta 17

## RESULT 9

US-09-817-879-6920  
; Sequence 6920, Application US/09817879  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals Inc.  
; TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Related  
; FILE REFERENCE: MH000-801-F  
; CURRENT APPLICATION NUMBER: US/09/817,879  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 9703  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6920  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: DNazyme  
US-09-817-879-6920

Query Match 61.0%; Score 12.2; DB 6; Length 31;  
Best Local Similarity 82.4%; Pred. No. 5.8e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttgctagatgctagta 19  
| | | | | | | | | |  
Db 1 ttgctagaggtagta 17

## RESULT 10

US-09-817-879-9054  
; Sequence 9054, Application US/09817879  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals Inc.  
; TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Related  
; FILE REFERENCE: MH000-801-F  
; CURRENT APPLICATION NUMBER: US/09/817,879  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 9703  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9054  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: DNazyme  
US-09-817-879-9054

Query Match 61.0%; Score 12.2; DB 6; Length 31;  
Best Local Similarity 82.4%; Pred. No. 5.8e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttgctagatgctagta 19  
| | | | | | | | | |  
Db 1 ttgctagaggtagta 17

## RESULT 11

US-60-253-378-11076/c  
; Sequence 11076, Application US/60253378  
; GENERAL INFORMATION:  
; APPLICANT: Havukkala, Ilkka J  
; TITLE OF INVENTION: Polynucleotides isolated from plants  
; FILE REFERENCE: and methods for their use.

; CURRENT APPLICATION NUMBER: US/60/253.378  
; CURRENT FILING DATE: 2000-11-27  
; NUMBER OF SEQ ID NOS: 40367  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11076  
; LENGTH: 36  
; TYPE: DNA  
; ORGANISM: Pinus radiata  
US-60-253-378-11076

Query Match 60.0%; Score 12; DB 8; Length 36;  
Best Local Similarity 75.0%; Pred. No. 7.5e+03;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ttgtgctagctaggtat 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 32 TTTGTCCTGTCGTAGGTAT 13

RESULT 12  
US-09-690-16613  
; Sequence 16613, Application US/09708690  
; GENERAL INFORMATION:-  
; APPLICANT: Ribozyme Pharmaceuticals, Inc.  
; APPLICANT: Pavco, Pam  
; APPLICANT: McSwiggen, Jim  
; APPLICANT: Stinchcomb, Dan  
; APPLICANT: Escobedo, Jaime  
; TITLE OF INVENTION: Method and Reagent for the Treatment of Diseases or Conditions Re  
; TITLE OF INVENTION: Levels of Vascular Endothelial Growth Factor Receptor  
; FILE REFERENCE: MBH00,876-L (400/002)  
; CURRENT APPLICATION NUMBER: US/09/708,690  
; CURRENT FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 60/005,974  
; PRIOR FILING DATE: 1995-10-26  
; PRIOR APPLICATION NUMBER: US 08/584,040  
; PRIOR FILING DATE: 1996-01-08  
; PRIOR APPLICATION NUMBER: US 09/371,772  
; PRIOR FILING DATE: 1999-08-10  
; PRIOR APPLICATION NUMBER: US 09/685,664  
; PRIOR FILING DATE: 2000-10-10  
; NUMBER OF SEQ ID NOS: 20828  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 16613  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Enzymatic Nucleic Acid  
US-09-708-690-16613

Query Match 59.0%; Score 11.8; DB 5; Length 31;  
Best Local Similarity 86.7%; Pred. No. 9.4e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 gctagatgctaggtta 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 3 gctagaggctagcta 17

RESULT 13  
US-09-232-170-20/c  
; Sequence 20, Application US/09232170  
; GENERAL INFORMATION:  
; APPLICANT: della-Cioppa, Guy  
; APPLICANT: Erwin, Robert L.  
; APPLICANT: Fitzmaurice, Wayne P.  
; APPLICANT: Hanley, Kathleen M.  
; APPLICANT: Kumagai, Monto H.  
; APPLICANT: Lindbo, John A.  
; APPLICANT: McGeer, David R.

; APPLICANT: Padgett, Hal S.  
; APPLICANT: Poque, Gregory P.  
; TITLE OF INVENTION: METHOD OF DETERMINING THE FUNCTION OF  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND THE PROTEINS THEY ENCODE BY  
; TITLE OF INVENTION: TRANSFECTING THE SAME INTO A HOST  
; FILE REFERENCE: 08010137US01  
; CURRENT APPLICATION NUMBER: US/09/232.170  
; CURRENT FILING DATE: 1999-01-15  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 20  
; LENGTH: 36  
; TYPE: DNA  
; ORGANISM: VIRAL  
US-09-232-170-20

Query Match 59.0%; Score 11.8; DB 6; Length 36;  
Best Local Similarity 86.7%; Pred. No. 9.6e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgtgctagctagcta 15  
| | | | | | | | | | | | | | | | | | | | | |  
Db 24 TCTTGCTAGATTCTA 10

RESULT 14  
US-09-840-424-773  
; Sequence 773, Application US/09840424  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: McCarthy, Sean A.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A  
; TITLE OF INVENTION: HUMAN MICROVASCULAR ENDOTHELIAL LIBRARY  
; FILE REFERENCE: 1600.1010-002  
; CURRENT APPLICATION NUMBER: US/09/840,424  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/084,098  
; PRIOR FILING DATE: 1998-05-04  
; PRIOR APPLICATION NUMBER: US 60/123,523  
; PRIOR FILING DATE: 1999-03-08  
; PRIOR APPLICATION NUMBER: US 60/126,974  
; PRIOR FILING DATE: 1999-03-30  
; PRIOR APPLICATION NUMBER: US 09/304,649  
; PRIOR FILING DATE: 1999-05-04  
; NUMBER OF SEQ ID NOS: 1311  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 773  
; LENGTH: 38  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(38)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-840-424-773

Query Match 59.0%; Score 11.8; DB 7; Length 38;  
Best Local Similarity 86.7%; Pred. No. 9.6e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 6 ctatagctaggtat 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 16 ctataggttagctat 30

RESULT 15  
US-09-801-274-1751  
; Sequence 1751, Application US/09801274  
; GENERAL INFORMATION:  
; APPLICANT: Cargill, Michele  
; APPLICANT: Ireland, James S.

; APPLICANT: Lander, Eric S.  
; TITLE OF INVENTION: HUMAN SINGLE NUCLEOTIDE POLYMORPHISMS  
; FILE REFERENCE: 2825.2009-001  
; CURRENT APPLICATION NUMBER: US/09/801,274  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 60/187,510  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: US 60/206,129  
; PRIOR FILING DATE: 2000-05-22  
; NUMBER OF SEQ ID NOS: 1802  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1751  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-801-274-1751

Query Match 58.0%; Score 11.6; DB 6; Length 31;  
Best Local Similarity 70.0%; Pred. No. 1.2e+04;  
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 ttgtgctagatgctaggat 20  
||| ||| ||| |||  
Db 8 ttctctasttgcaatgat 27

Search completed: October 2, 2001, 16:55:01  
Job time: 17664 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 21:50:10 ; Search time 17695.9 Seconds  
(without alignments)  
16.701 Million cell updates/sec

Title: US-09-757-100B-8

Perfect score: 20

Sequence: 1 ctatgctgtagtctgtc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1415048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 4519004

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:\*

- 1: /cgn2\_6/ptodata/1/pna/US0612\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pna/US0613\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pna/US08\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq:\*
- 20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq:\*
- 21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq:\*
- 22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq:\*
- 23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq:\*
- 24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq:\*
- 25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq:\*
- 26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq:\*
- 27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq:\*
- 28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq:\*
- 29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq:\*
- 30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq:\*
- 31: /cgn2\_6/ptodata/1/pna/US098\_COMB.seq:\*
- 32: /cgn2\_6/ptodata/1/pna/US0981\_COMB.seq:\*
- 33: /cgn2\_6/ptodata/1/pna/US0982\_COMB.seq:\*
- 34: /cgn2\_6/ptodata/1/pna/US0983\_COMB.seq:\*
- 35: /cgn2\_6/ptodata/1/pna/US0984\_COMB.seq:\*
- 36: /cgn2\_6/ptodata/1/pna/US0985\_COMB.seq:\*
- 37: /cgn2\_6/ptodata/1/pna/US0986\_COMB.seq:\*
- 38: /cgn2\_6/ptodata/1/pna/US0987\_COMB.seq:\*
- 39: /cgn2\_6/ptodata/1/pna/US0988\_COMB.seq:\*
- 40: /cgn2\_6/ptodata/1/pna/US0989\_COMB.seq:\*
- 41: /cgn2\_6/ptodata/1/pna/US0990\_COMB.seq:\*
- 42: /cgn2\_6/ptodata/1/pna/US0991\_COMB.seq:\*
- 43: /cgn2\_6/ptodata/1/pna/US0992\_COMB.seq:\*

- 44: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq:\*
- 45: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq:\*
- 46: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq:\*
- 47: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq:\*
- 48: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq:\*
- 49: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq:\*
- 50: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq:\*
- 51: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq:\*
- 52: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq:\*
- 53: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq:\*
- 54: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq:\*
- 55: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq:\*
- 56: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq:\*
- 57: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq:\*
- 58: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq:\*
- 59: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq:\*
- 60: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	20	1	PCT-US00-18999-8
2	20	100.0	20	17	US-09-377-310-8
3	20	100.0	20	29	US-09-757-100B-8
4	15.8	79.0	25	55	US-60-233-620-73486
5	15.8	79.0	25	55	US-60-233-620-73499
6	15.4	77.0	47	53	US-60-216-745-2764
7	15.2	76.0	25	55	US-60-233-620-101590
8	15	75.0	15	1	PCT-US00-18999-28
9	15	75.0	15	17	US-09-377-310-28
10	15	75.0	15	29	US-09-757-100B-28
11	15	75.0	20	1	PCT-US00-18999-9
12	15	75.0	20	17	US-09-377-310-9
13	15	75.0	20	29	US-09-757-100B-9
14	14.2	71.0	25	26	US-09-660-220-11883
15	14.2	71.0	25	26	US-09-660-220-48863
16	13.8	69.0	25	55	US-60-232-638-99167
17	13.6	68.0	21	18	US-09-422-978-8159
18	13.6	68.0	25	55	US-60-233-166-175158
19	13.6	68.0	25	55	US-60-233-166-218150
20	13.6	68.0	25	55	US-60-233-166-235863
21	13.6	68.0	25	55	US-60-233-166-321589
22	13.6	68.0	25	55	US-60-233-166-89443
23	13.4	67.0	25	55	US-60-234-017-94831
24	13.4	67.0	43	3	US-07-858-747A-17
25	13.4	67.0	43	18	US-09-414-117-17
26	13.2	66.0	25	26	US-09-660-220-48858
27	13.2	66.0	25	55	US-60-232-638-71961
28	13.2	66.0	25	55	US-60-232-638-123715
29	13.2	66.0	25	55	US-60-233-166-276832
30	13.2	66.0	25	55	US-60-233-166-359242
31	13.2	66.0	25	55	US-60-233-620-38163
32	13.2	66.0	25	55	US-60-233-620-65887
33	13.2	66.0	25	55	US-60-233-620-65897
34	13.2	66.0	25	55	US-60-234-017-4607
35	13.2	66.0	25	55	US-60-234-017-222942
36	13.2	66.0	25	55	US-60-234-017-458673
37	13.2	66.0	25	55	US-60-234-017-458675
38	13.2	66.0	50	13	US-08-900-106-532
39	13	65.0	25	55	US-60-232-638-58459
40	12.8	64.0	24	14	US-09-068-416-5
41	12.8	64.0	25	26	US-09-660-220-14687
42	12.8	64.0	25	26	US-09-660-220-68406
43	12.8	64.0	25	55	US-60-233-620-13600
44	12.8	64.0	25	55	US-60-233-620-13607
45	12.8	64.0	25	55	US-60-233-620-13608

## ALIGNMENTS

RESULT 1  
PCT-US00-18999-8  
; Sequence 8, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0476  
; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-8

Query Match 100.0%; Score 20; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ctatagctaggtatctgtc 20

RESULT 2  
US-09-377-310-8  
; Sequence 8, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-8

Query Match 100.0%; Score 20; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ctatagctaggtatctgtc 20

RESULT 3  
US-09-757-100B-8  
; Sequence 8, Application US/09757100B

; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100B-8

Query Match 100.0%; Score 20; DB 29; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ctatagctaggtatctgtc 20

RESULT 4  
US-60-233-620-73486  
; Sequence 73486, Application US/60233620  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of  
; FILE REFERENCE: 3116  
; CURRENT APPLICATION NUMBER: US/60/233,620  
; CURRENT FILING DATE: 2000-10-24  
; NUMBER OF SEQ ID NOS: 131820  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 73486  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AC003672  
US-60-233-620-73486

Query Match 79.0%; Score 15.8; DB 55; Length 25;  
Best Local Similarity 89.5%; Pred. No. 3.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tagatgctaggtatctgtc 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 4 tagaggctaggtatcttctc 22

RESULT 5  
US-60-233-620-73499  
; Sequence 73499, Application US/60233620  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of  
; FILE REFERENCE: 3116  
US-60-233-620-73499



### Query Match

Best Local Similarity 100.0%; Pred. No. 8.2e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 agatgctaggtagtct 17  
|||||  
Db 1 agatgctaggtagtct 15

RESULT 10  
US-09-757-100b-28  
; Sequence 28, Application US/09757100B  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR FILING DATE: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100b-28

Query Match 75.0%; Score 15; DB 29; Length 15;  
Best Local Similarity 100.0%; Pred. No. 8.2e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 agatgctaggtagtct 17  
|||||  
Db 1 agatgctaggtagtct 15

RESULT 11  
PCT-US00-18999-9  
; Sequence 9, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0476  
; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; CURRENT FILING DATE: 2000-07-13  
; PRIOR FILING DATE: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-9

Query Match 75.0%; Score 15; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.6e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatagtgctaggtagt 15  
|||||  
Db 6 ctatagtgctaggtagt 20

RESULT 12  
US-09-377-310-9  
; Sequence 9, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-9

Query Match 75.0%; Score 15; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.6e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatagtgctaggtagt 15  
|||||  
Db 6 ctatagtgctaggtagt 20

RESULT 13  
US-09-757-100b-9  
; Sequence 9, Application US/09757100B  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR FILING DATE: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100b-9

Query Match 75.0%; Score 15; DB 29; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.6e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatagtgctaggtagt 15  
|||||  
Db 6 ctatagtgctaggtagt 20

## RESULT 14

US-09-660-220-11883  
; Sequence 11883, Application US/09660220  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann et al.  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of Human  
; FILE REFERENCE: 3102.1  
; CURRENT APPLICATION NUMBER: US/09/660,220  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 60/164,973  
; PRIOR FILING DATE: 1999-11-11  
; NUMBER OF SEQ ID NOS: 140981  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11883  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank D63881  
US-09-660-220-11883

Query Match 71.0%; Score 14.2; DB 26; Length 25;  
Best Local Similarity 84.2%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ctatgctcctagggtatctgt 19  
| | | | | | | | | | | | | | | | | | | | |  
Db 2 ctatgctcctagggtatctgt 20

## RESULT 15

US-09-660-220-48863/c  
; Sequence 48863, Application US/09660220  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann et al.  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of Human  
; FILE REFERENCE: 3102.1  
; CURRENT APPLICATION NUMBER: US/09/660,220  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 60/164,973  
; PRIOR FILING DATE: 1999-11-11  
; NUMBER OF SEQ ID NOS: 140981  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48863  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank M14338  
US-09-660-220-48863

Query Match 71.0%; Score 14.2; DB 26; Length 25;  
Best Local Similarity 84.2%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tagatgctcctagggtatctgtc 20  
| | | | | | | | | | | | | | | | | | | | |  
Db 24 TAGATGCTCCTAGGGTATCTGTC 6

Search completed: October 2, 2001, 21:50:11  
Job time: 24524 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:55:00 ; Search time 2173.58 Seconds  
(without alignments)  
19.290 Million cell updates/sec

Title: US-09-757-100B-8

Perfect score: 20

Sequence: 1 ctatagctaggtatctgtc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1895189 seqs, 1048201267 residues

Total number of hits satisfying chosen parameters: 299022

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_NA\_New.\*  
1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	12.8	64.0	38	7	US-09-840-424-773
2	12.2	61.0	45	8	US-60-278-258-6107
3	12	60.0	17	6	US-09-546-745A-3190
4	12	60.0	17	6	US-09-546-745A-3191
5	12	60.0	36	7	US-09-048-473-3
6	12	60.0	36	7	US-09-048-473-7
7	12	60.0	39	8	US-60-253-457-43320
8	11.8	59.0	17	6	US-09-866-108-2604
9	11.8	59.0	17	6	US-09-866-108-2605
10	11.8	59.0	17	6	US-09-866-108-2606
11	11.8	59.0	25	6	US-09-866-108-5533
12	11.8	59.0	25	6	US-09-866-108-5534
13	11.8	59.0	25	6	US-09-866-108-5535
14	11.8	59.0	25	6	US-09-866-108-5536
15	11.8	59.0	25	6	US-09-866-108-5537
16	11.8	59.0	25	6	US-09-866-108-5538
17	11.8	59.0	25	6	US-09-866-108-5539
18	11.8	59.0	25	6	US-09-866-108-5540
19	11.8	59.0	25	6	US-09-866-108-5541
20	11.8	59.0	25	6	US-09-866-108-5542
21	11.8	59.0	25	6	US-09-866-108-5543
22	11.8	59.0	50	4	US-08-798-074B-7431
23	11.8	59.0	50	4	US-08-798-074B-7431
24	11.8	59.0	50	4	US-08-798-074C-7431
25	11.8	59.0	50	7	US-09-839-976-3279

Sequence 1033, Ap  
Sequence 8033, Ap  
Sequence 3189, Ap  
Sequence 3, Appl  
Sequence 32, Appl  
Sequence 10, Appl  
Sequence 26932, A  
Sequence 87, Appl  
Sequence 191, App  
Sequence 22107, A  
Sequence 16933, A  
Sequence 17980, A  
Sequence 18029, A  
Sequence 2187, Ap  
Sequence 34, Appl  
Sequence 235, App  
Sequence 3466, Ap  
Sequence 2603, Ap  
Sequence 2607, Ap  
Sequence 100, App

US-09-825-790-1033  
US-09-708-690-8033  
US-09-546-745A-3189  
US-09-625-225-3  
US-09-801-274-32  
US-09-545-551A-10  
US-60-253-651-26932  
US-09-917-265-87  
US-09-827-383-191  
US-09-724-750-22107  
US-09-708-690-16933  
US-09-708-690-17980  
US-09-708-690-18029  
US-09-864-785-2187  
US-09-743-282-34  
US-09-724-671-235  
US-60-278-561-3466  
US-09-866-108-2603  
US-09-866-108-2607  
US-09-635-251-100

#### ALIGNMENTS

RESULT 1

US-09-840-424-773

; Sequence 773, Application US/09840424

; GENERAL INFORMATION:

; APPLICANT: Gearing, David P.

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A

; FILE REFERENCE: 1600.1010-002

; CURRENT APPLICATION NUMBER: US/09/840.424

; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: US 60/084,098

; PRIOR FILING DATE: 1998-05-04

; PRIOR APPLICATION NUMBER: US 60/123,523

; PRIOR FILING DATE: 1999-03-08

; PRIOR APPLICATION NUMBER: US 60/126,974

; PRIOR FILING DATE: 1999-03-30

; PRIOR APPLICATION NUMBER: US 09/304,649

; PRIOR FILING DATE: 1999-05-04

; NUMBER OF SEQ ID NOS: 1311

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 773

; LENGTH: 38

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(38)

; OTHER INFORMATION: n = A,T,C or G

US-09-840-424-773

Query Match 64.0%; Score 12.8; DB 7; Length 38;

Best Local Similarity 87.5%; Pred. No. 1.5e+03;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctatagctaggtatc 16

||||||| || |||||

Db 16 ctatagtttaggtatc 31

RESULT 2

US-60-278-258-6107/c

; Sequence 6107, Application US/60278258

; GENERAL INFORMATION:

; APPLICANT: Morris, MacDonald

; APPLICANT: Lal, Preeti

; APPLICANT: Diep, Dinh

; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using  
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide  
; FILE REFERENCE: GX-0010-1.P  
; CURRENT APPLICATION NUMBER: US/60/278,258  
; NUMBER OF SEQ ID NOS: 2001-03-23  
; SOFTWARE: PERL Program  
; SEQ ID NO 6107  
; LENGTH: 45  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: SNP00057112  
; NAME/KEY: snp  
; LOCATION: 20  
; OTHER INFORMATION: 350403.1, 20, G->A  
US-60-278-258-6107

Query Match 61.0%; Score 12.2; DB 8; Length 45;  
Best Local Similarity 82.4%; Pred. No. 3.5e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ctatgctgctaggtatct 17  
||||| ||| ||| |  
Db 18 CTAGATGGTAGTATGCT 2

RESULT 3  
US-09-546-745A-3190  
; Sequence 3190, Application US/09546745A  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals, Inc.  
; APPLICANT: Blatt, Larry  
; APPLICANT: Zwick, Michael  
; APPLICANT: Pavco, Pam  
; APPLICANT: McSwiggen, Jim  
; TITLE OF INVENTION: Regulation of Repressor Genes using Nucleic Acid Molecules  
; FILE REFERENCE: 237/193  
; CURRENT APPLICATION NUMBER: US/09/546,745A  
; CURRENT FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 7043  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3190  
; LENGTH: 17  
; TYPE: RNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic Acid  
US-09-546-745A-3190

Query Match 60.0%; Score 12; DB 6; Length 17;  
Best Local Similarity 58.3%; Pred. No. 3.8e+03;  
Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 8 ctagggtatctgt 19  
|:||||:|:|:  
Db 3 cuagguaucugu 14

RESULT 4  
US-09-546-745A-3191  
; Sequence 3191, Application US/09546745A  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals, Inc.  
; APPLICANT: Blatt, Larry  
; APPLICANT: Zwick, Michael  
; APPLICANT: Pavco, Pam  
; APPLICANT: McSwiggen, Jim  
; TITLE OF INVENTION: Regulation of Repressor Genes using Nucleic Acid Molecules

; FILE REFERENCE: 237/193  
; CURRENT APPLICATION NUMBER: US/09/546,745A  
; CURRENT FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 7043  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3191  
; LENGTH: 17  
; TYPE: RNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic Acid  
US-09-546-745A-3191

Query Match 60.0%; Score 12; DB 6; Length 17;  
Best Local Similarity 58.3%; Pred. No. 3.8e+03;  
Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 8 ctagggtatctgt 19  
|:||||:|:|:  
Db 1 cuagguaucugu 12

RESULT 5  
US-09-048-473-3/c  
; Sequence 3, Application US/09048473  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; Hayenga, Kirk J.  
; Lawlis, Virgil B.  
; Ward, Michael  
; TITLE OF INVENTION: ASPARTIC PROTEINASE DEFICIENT  
FILAMENTOUS FUNGI  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International, Inc.  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/048,473  
FILING DATE: 26-Mar-1998  
; PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/214,237  
FILING DATE: 01-JUL-1992  
APPLICATION NUMBER: 07/931,123  
FILING DATE: 17-AUG-1992  
APPLICATION NUMBER: 08/345,018  
FILING DATE: 23-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J.  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC45-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-6504  
FAX: 650-846-7620  
; INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-048-473-3

Query Match 60.0%; Score 12; DB 7; Length 36;  
Best Local Similarity 75.0%; Pred. No. 4.3e+03;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20  
||||| | ||| ||| |||  
Db 22 CTAGACTCGAGGATCCGTC 3

## RESULT 6

US-09-048-473-7/c

; Sequence 7, Application US/09048473

; GENERAL INFORMATION:

; APPLICANT: Berkka, Randy M.

; Hayenga, Kirk J.

; Lawlis, Virgil B.

; Ward, Michael

; TITLE OF INVENTION: ASPARTIC PROTEINASE DEFICIENT

; FILAMENTOUS FUNGI

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genencor International, Inc.

; STREET: 925 Page Mill Road

; CITY: Palo Alto

; STATE: CA

; COUNTRY: US

; ZIP: 94304-1013

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/048,473

; FILING DATE: 26-Mar-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/214,237

; FILING DATE: 01-JUL-1992

; APPLICATION NUMBER: 07/931,123

; FILING DATE: 17-AUG-1992

; APPLICATION NUMBER: 08/345,018

; FILING DATE: 23-NOV-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Glaister, Debra J.

; REGISTRATION NUMBER: 33,888

; REFERENCE/DOCKET NUMBER: GC45-4

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-846-7620

; TELEFAX: 650-845-6504

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 36 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-09-048-473-7

Query Match 60.0%; Score 12; DB 7; Length 36;  
Best Local Similarity 75.0%; Pred. No. 4.3e+03;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20  
||||| | ||| ||| |||  
Db 22 CTAGACTCGAGGATCCGTC 3

## RESULT 7

US-60-253-457-43320/c

; Sequence 43320, Application US/60253457

; GENERAL INFORMATION:

; APPLICANT: Havukkala, Ilkka J

; TITLE OF INVENTION: Polynucleotides, isolated from plants  
; TITLE OF INVENTION: and methods for their use.

; FILE REFERENCE: 1054P2

; CURRENT APPLICATION NUMBER: US/60/253,457

; CURRENT FILING DATE: 2000-11-27

; NUMBER OF SEQ ID NOS: 48893

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 43320

; LENGTH: 39

; TYPE: DNA

; ORGANISM: Pinus radiata

US-60-253-457-43320

Query Match 60.0%; Score 12; DB 8; Length 39;

Best Local Similarity 75.0%; Pred. No. 4.4e+03;

Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20  
||||| | ||| ||| |||  
Db 22 CTCGAGGCTAGTCTCTCTC 3

## RESULT 8

US-09-866-108-2604

; Sequence 2604, Application US/09866108

; GENERAL INFORMATION:

; APPLICANT: GU, Yizhong

; APPLICANT: JI, Yonggang

; APPLICANT: PENN, Sharron G.

; APPLICANT: HANZEL, David K.

; APPLICANT: RANK, David R.

; APPLICANT: CHEN, Wensheng

; APPLICANT: SHANNON, Mark

; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE

; FILE REFERENCE: AEOMICA-7

; CURRENT APPLICATION NUMBER: US/09/866,108

; CURRENT FILING DATE: 2001-05-25

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 60/266,860

; PRIOR FILING DATE: 2001-02-05

; NUMBER OF SEQ ID NOS: 15752

; SOFTWARE: Acomica Sequence Listing Engine

; SEQ ID NO 2604

; LENGTH: 17

; TYPE: DNA

; ORGANISM: Homo sapiens  
US-09-866-108-2604

Query Match 59.0%; Score 11.8; DB 6; Length 17;  
Best Local Similarity 86.7%; Pred. No. 4.9e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agatgctagggtatct 17  
||||| |||

Db 3 agatgctgggcattct 17  
||||| |||

## RESULT 9

US-09-866-108-2605  
; Sequence 2605, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.  
; APPLICANT: HANZEL, David K.  
; APPLICANT: RANK, David R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: SHANNON, Mark  
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
; FILE REFERENCE: AEOMICA-7  
; CURRENT APPLICATION NUMBER: US/09/866,108  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 60/266,860  
; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aecomica Sequence Listing Engine  
; SEQ ID NO 2605  
; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-2605

Query Match 59.0%; Score 11.8; DB 6; Length 17;  
Best Local Similarity 86.7%; Pred. No. 4.9e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agatgctagggtatct 17  
||||| |||

Db 2 agatgctgggcattct 16

## RESULT 10

US-09-866-108-2606  
; Sequence 2606, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.  
; APPLICANT: HANZEL, David K.  
; APPLICANT: RANK, David R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: SHANNON, Mark  
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
; FILE REFERENCE: AEOMICA-7  
; CURRENT APPLICATION NUMBER: US/09/866,108  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 60/266,860  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aecomica Sequence Listing Engine  
; SEQ ID NO 2606  
; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-2606

Query Match 59.0%; Score 11.8; DB 6; Length 17;  
Best Local Similarity 86.7%; Pred. No. 4.9e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agatgctagggtatct 17  
||||| |||

Db 1 agatgctgggcattct 15

## RESULT 11

US-09-866-108-5533  
; Sequence 5533, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.





```
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aecomica Sequence Listing Engine
; SEQ ID NO 5535
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-5535
```

```
Query Match 59.0%; Score 11.8; DB 6; Length 25;
Best Local Similarity 86.7%; Pred. No. 5.3e+03;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 3 agatgctaggtatct 17
   ||||| |||||
Db 9 agatgctggcatct 23
```

## RESULT 14

```
US-09-866-108-5536
; Sequence 5536, Application US/09866108
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aecomica Sequence Listing Engine
; SEQ ID NO 5536
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-5536
```

```
Query Match 59.0%; Score 11.8; DB 6; Length 25;
Best Local Similarity 86.7%; Pred. No. 5.3e+03;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 3 agatgctaggtatct 17
   ||||| |||||
Db 8 agatgctggcatct 22
```

## RESULT 15

```
US-09-866-108-5537
; Sequence 5537, Application US/09866108
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aecomica Sequence Listing Engine
```

; SEQ ID NO 5537  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-5537

Query Match 59.0%; Score 11.8; DB 6; Length 25;  
Best Local Similarity 86.7%; Pred. No. 5.3e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agatgctaggtatct 17  
||||| |||||  
Db 7 agatgctgggcattct 21

Search completed: October 2, 2001, 16:55:01  
Job time: 17664 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 21:50:09 ; Search time 17695.9 Seconds  
(without alignments)  
16.701 Million cell updates/sec

Title: US-09-757-100B-7

Perfect score: 20

Sequence: 1 ttctcccttcggtattctt 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 4519004

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_NA\_Main:\*

1: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq.\*

2: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq.\*

3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*

4: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*

5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq.\*

6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq.\*

7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq.\*

8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq.\*

9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*

10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*

11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq.\*

12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq.\*

13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq.\*

14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq.\*

15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq.\*

16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq.\*

17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq.\*

18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq.\*

19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*

20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*

21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*

22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*

23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*

24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*

25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*

26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq.\*

27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*

28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*

29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*

30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*

31: /cgn2\_6/ptodata/1/pna/US098\_COMB.seq.\*

32: /cgn2\_6/ptodata/1/pna/US099\_COMB.seq.\*

33: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*

34: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*

35: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*

36: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*

37: /cgn2\_6/ptodata/1/pna/US099E\_COMB.seq.\*

38: /cgn2\_6/ptodata/1/pna/US099F\_COMB.seq.\*

39: /cgn2\_6/ptodata/1/pna/US099G\_COMB.seq.\*

40: /cgn2\_6/ptodata/1/pna/US099H\_COMB.seq.\*

41: /cgn2\_6/ptodata/1/pna/US099I\_COMB.seq.\*

42: /cgn2\_6/ptodata/1/pna/US099J\_COMB.seq.\*

43: /cgn2\_6/ptodata/1/pna/US099K\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*

45: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*

46: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq.\*

47: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*

48: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*

49: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*

50: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq.\*

51: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq.\*

52: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq.\*

53: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq.\*

54: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq.\*

55: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq.\*

56: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq.\*

57: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq.\*

58: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq.\*

59: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq.\*

60: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	20	1	PCT-US00-18999-7
2	20	100.0	20	17	US-09-377-310-7
3	20	100.0	20	29	US-09-757-100B-7
C	15.4	77.0	44	17	US-09-310-298-486
5	15.2	76.0	36	18	US-09-403-752A-115
6	15	75.0	15	1	PCT-US00-18999-27
7	15	75.0	15	17	US-09-377-310-27
8	15	75.0	15	29	US-09-757-100B-27
9	14.8	74.0	25	55	US-60-233-166-23007
10	14.8	74.0	25	55	US-60-234-017-131414
11	14.2	71.0	25	55	US-60-233-166-280120
12	13.8	69.0	20	25	US-09-657-481A-39
C	13.8	69.0	25	55	US-60-233-166-108604
14	13.8	69.0	25	55	US-60-233-166-167371
15	13.8	69.0	25	55	US-60-234-017-425796
C	13.8	69.0	25	55	US-60-234-017-425802
16	13.8	69.0	33	14	US-09-016-061-83
18	13.8	69.0	33	17	US-09-339-922A-83
19	13.6	68.0	25	55	US-60-232-638-106175
20	13.6	68.0	25	55	US-60-233-166-168611
C	13.6	68.0	25	55	US-60-233-166-330256
22	13.6	68.0	50	12	US-08-899-858-2789
C	13.6	68.0	50	34	US-60-023-277-2789
24	13.4	67.0	25	55	US-60-233-620-25074
25	13.2	66.0	18	53	US-60-216-745-6364
C	13.2	66.0	20	18	US-09-422-978-8164
26	13.2	66.0	21	53	US-60-216-745-5644
C	13.2	66.0	24	1	PCT-US00-24784-163
28	13.2	66.0	25	26	US-09-660-220-87881
29	13.2	66.0	25	55	US-60-232-638-66944
30	13.2	66.0	25	55	US-60-232-638-70508
31	13.2	66.0	25	55	US-60-232-638-85391
C	13.2	66.0	25	55	US-60-233-166-26780
33	13.2	66.0	25	55	US-60-233-166-205876
34	13.2	66.0	25	55	US-60-233-166-333957
35	13.2	66.0	25	55	US-60-233-166-420026
36	13.2	66.0	25	55	US-60-234-017-43611
37	13.2	66.0	25	55	US-60-234-017-287051
C	13.2	66.0	25	55	US-60-234-017-297294
42	13.2	66.0	25	55	US-60-234-017-330691
43	13.2	66.0	25	55	US-60-234-017-528222
44	13.2	66.0	25	55	US-60-234-049-129873
C	13.2	66.0	25	55	US-60-234-049-129873

## ALIGNMENTS

RESULT 1  
PCT-US00-18999-7  
; Sequence 7, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0476  
; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; CURRENT FILING DATE: 2000-07-13  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-7

Query Match 100.0%; Score 20; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
Db 1 ttctcccttcggtattctt 20

RESULT 2  
US-09-377-310-7  
; Sequence 7, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-7

Query Match 100.0%; Score 20; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
Db 1 ttctcccttcggtattctt 20

RESULT 3  
US-09-757-100b-7  
; Sequence 7, Application US/09757100B

; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100b-7

Query Match 100.0%; Score 20; DB 29; Length 20;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
Db 1 ttctcccttcggtattctt 20

RESULT 4  
US-09-310-298-486/c  
; Sequence 486, Application US/09310298  
; GENERAL INFORMATION:  
; APPLICANT: Lander, Eric S.  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: PRIMER SEQUENCES FOR SINGLE NUCLEOTIDE  
; FILE REFERENCE: WHIFG98-07pA  
; CURRENT APPLICATION NUMBER: US/09/310,298  
; CURRENT FILING DATE: 1998-05-12  
; EARLIER APPLICATION NUMBER: 60/085,152  
; EARLIER FILING DATE: 1998-05-12  
; NUMBER OF SEQ ID NOS: 5968  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 486  
; LENGTH: 44  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide primer  
US-09-310-298-486

Query Match 77.0%; Score 15.4; DB 17; Length 44;  
Best Local Similarity 94.1%; Pred. No. 2.7e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 tcccttcggtattctt 20  
Db 39 TCCCTTCGTTATTCAT 23

RESULT 5  
US-09-403-752A-115  
; Sequence 115, Application US/09403752A  
; GENERAL INFORMATION:  
; APPLICANT: Borgford, Thor  
; TITLE OF INVENTION: RICIN-LIKE TOXIN VARIANTS FOR TREATMENT OF CANCER,  
; TITLE OF INVENTION: VIRAL OR PARASITIC INFECTIONS

```
; FILE REFERENCE: 10447-005
; CURRENT APPLICATION NUMBER: US/09/403,752A
; CURRENT FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: U.S. 60/045,148
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: U.S. 60/063,715
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 142
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 115
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Human Prostate-Specific Antigen linker region of PAP-290
US-09-403-752A-115

Query Match          76.0%; Score 15.2; DB 18; Length 36;
Best Local Similarity 85.0%; Pred. No. 3.3e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttctccttcgcgttattctt 20
Db 14 ttctccttcgcgatattttt 33

RESULT 6
PCT-US00-18999-27
; Sequence 27, Application PC/TUS0018999
; GENERAL INFORMATION:
; APPLICANT: Isis Pharmaceuticals, Inc.
; APPLICANT: Monia, Brett P.
; APPLICANT: Gaarde, William A.
; APPLICANT: Nero, Pamela S.
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase
; TITLE OF INVENTION: Expression
; FILE REFERENCE: ISPH-0476
; CURRENT APPLICATION NUMBER: PCT/US00/18999
; CURRENT FILING DATE: 2000-07-13
; PRIOR APPLICATION NUMBER: 09/377,310
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: antisense sequence
PCT-US00-18999-27

Query Match          75.0%; Score 15; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.5e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ctcccttcgcgttatt 17
Db 1 ctcccttcgcgttatt 15

RESULT 7
US-09-377-310-27
; Sequence 27, Application US/09377310A
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Gaarde, William A.
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase
; TITLE OF INVENTION: Expression
; FILE REFERENCE: ISPH-0389
; CURRENT APPLICATION NUMBER: US/09/377,310A
; CURRENT FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 27
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: antisense sequence
US-09-377-310-27

Query Match          75.0%; Score 15; DB 17; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.5e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ctcccttcgcgttatt 17
Db 1 ctcccttcgcgttatt 15

RESULT 8
US-09-757-100B-27
; Sequence 27, Application US/09757100B
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Gaarde, William A.
; APPLICANT: Nero, Pamela S.
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase
; TITLE OF INVENTION: Expression
; FILE REFERENCE: ISPH-0533
; CURRENT APPLICATION NUMBER: US/09/757,100B
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 09/377,310
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/US00/18999
; PRIOR FILING DATE: 2000-07-13
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: antisense sequence
US-09-757-100B-27

Query Match          75.0%; Score 15; DB 29; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.5e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ctcccttcgcgttatt 17
Db 1 ctcccttcgcgttatt 15

RESULT 9
US-60-233-166-23007
; Sequence 23007, Application US/60233166
; GENERAL INFORMATION:
; APPLICANT: Mitmann
; APPLICANT: Affymetrix, Inc.
; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 3112
; CURRENT APPLICATION NUMBER: US/60/233,166
; CURRENT FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 420907
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23007
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank AA818353
```

## US-60-233-166-23007

Query Match 74.0%; Score 14.8; DB 55; Length 25;  
Best Local Similarity 88.9%; Pred. No. 4.8e+03;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 ctccctccggtattctt 20  
||||| ||| |||

Db 8 ctccctccggtattctt 25

## RESULT 10

US-60-234-017-131414  
; Sequence 131414, Application US/60234017

; GENERAL INFORMATION:

; APPLICANT: Mittmann, M

; APPLICANT: Affymetrix, Inc.

; TITLE OF INVENTION: Methods of Genetic Analysis of Mus

; FILE REFERENCE: 3115

; CURRENT APPLICATION NUMBER: US/60/234,017

; CURRENT FILING DATE: 2000-09-20

; NUMBER OF SEQ ID NOS: 605887

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 131414

; LENGTH: 25

; TYPE: DNA

; ORGANISM: Mus musculus

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: GenBank AA600645

US-60-234-017-131414

Query Match 74.0%; Score 14.8; DB 55; Length 25;  
Best Local Similarity 88.9%; Pred. No. 4.8e+03;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 tctccctccggtattctt 19  
|| || ||| ||| ||| |||

Db 8 tcaccttccggtattctt 25

## RESULT 11

US-60-233-166-280120  
; Sequence 280120, Application US/60233166

; GENERAL INFORMATION:

; APPLICANT: Mittmann

; APPLICANT: Affymetrix, Inc.

; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat

; FILE REFERENCE: 3112

; CURRENT APPLICATION NUMBER: US/60/233,166

; CURRENT FILING DATE: 2000-10-24

; NUMBER OF SEQ ID NOS: 420907

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 280120

; LENGTH: 25

; TYPE: DNA

; ORGANISM: Rattus norvegicus

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: GenBank AF639138

US-60-233-166-280120

Query Match 71.0%; Score 14.2; DB 55; Length 25;  
Best Local Similarity 84.2%; Pred. No. 9.2e+03;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ttctccctccggtattctt 19  
||| ||| ||| ||| ||| |||

Db 1 ttgcgacctccggtattctt 19

## RESULT 12

US-09-657-481A-39/c  
; Sequence 39, Application US/09657481A

; GENERAL INFORMATION:

; APPLICANT: Brett P. Monia

; APPLICANT: Lex M. Cowser

; TITLE OF INVENTION: ANTISENSE MODULATION OF UBIQUITIN PROTEIN LIGASE WWP1 AND

; FILE REFERENCE: RTS-0087

; CURRENT APPLICATION NUMBER: US/09/657,481A

; CURRENT FILING DATE: 2000-09-07

; NUMBER OF SEQ ID NOS: 93

; SEQ ID NO 39

; LENGTH: 20

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Antisense Oligonucleotide

US-09-657-481A-39

Query Match 69.0%; Score 13.8; DB 25; Length 20;  
Best Local Similarity 88.2%; Pred. No. 1.4e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ttctccctccggtattctt 17  
| ||| ||| ||| ||| |||

Db 18 TGCCTCACTCCGTTATT 2

## RESULT 13

US-60-233-166-108604/c  
; Sequence 108604, Application US/60233166

; GENERAL INFORMATION:

; APPLICANT: Mittmann

; APPLICANT: Affymetrix, Inc.

; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat

; FILE REFERENCE: 3112

; CURRENT APPLICATION NUMBER: US/60/233,166

; CURRENT FILING DATE: 2000-10-24

; NUMBER OF SEQ ID NOS: 420907

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 108604

; LENGTH: 25

; TYPE: DNA

; ORGANISM: Rattus norvegicus

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: GenBank AA860043

US-60-233-166-108604

Query Match 69.0%; Score 13.8; DB 55; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.4e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 ctccctccggtattctt 19  
||||| ||| ||| ||| |||

Db 21 CTCCTACCGTTGTCT 5

## RESULT 14

US-60-233-166-167371  
; Sequence 167371, Application US/60233166

; GENERAL INFORMATION:

; APPLICANT: Mittmann

; APPLICANT: Affymetrix, Inc.

; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat

; FILE REFERENCE: 3112

; CURRENT APPLICATION NUMBER: US/60/233,166

; CURRENT FILING DATE: 2000-10-24



; NUMBER OF SEQ ID NOS: 420907  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 167371  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AA926247  
US-60-233-166-167371

Query Match 69.0%; Score 13.8; DB 55; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.4e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tcccttcggttattctt 20  
||||| ||||| |||||  
Db 5 tcccttcggttattctt 21

RESULT 15  
US-60-234-017-425796/c  
; Sequence 425796, Application US/60234017  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann, M  
; TITLE OF INVENTION: Methods of Genetic Analysis of Mus  
; FILE REFERENCE: 3115  
; CURRENT APPLICATION NUMBER: US/60/234,017  
; CURRENT FILING DATE: 2000-09-20  
; NUMBER OF SEQ ID NOS: 605887  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 425796  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AV102460  
US-60-234-017-425796

Query Match 69.0%; Score 13.8; DB 55; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.4e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tctcccttcggttattc 18  
||||| ||||| |||||  
Db 18 TCTCCCTTCGCTAATC 2

Search completed: October 2, 2001, 21:50:10  
Job time: 24523 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:55:00 ; Search time 2173.58 Seconds  
(without alignments)  
19.290 Million cell updates/sec

Title: US-09-757-100B-7

Perfect score: 20

Sequence: 1 ttctcccttcggtattctt 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1895189 seqs, 1048201267 residues

Total number of hits satisfying chosen parameters: 299022

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_New.\*

- 1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 8: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	15.2	76.0	36	7	US-09-551-151-115
2	13.8	69.0	21	6	US-09-726-774-70
3	13.8	69.0	33	6	US-09-900-590-83
4	13.6	68.0	36	8	US-60-252-833-32245
5	12.6	63.0	30	4	US-08-850-062A-75
6	12.6	63.0	46	8	US-60-278-561-657
7	12.2	61.0	47	6	US-09-516-667-73
8	12.2	61.0	48	6	US-09-516-667-35
9	12.2	61.0	48	6	US-09-516-667-36
10	12.2	61.0	48	6	US-09-516-667-74
11	12.2	61.0	48	6	US-09-516-667-75
12	12.2	61.0	48	6	US-09-516-667-75
13	11.8	59.0	28	8	US-60-252-833-9824
14	11.6	58.0	29	8	US-60-253-652-14008
15	11.6	58.0	31	5	US-09-574-376B-90
16	11.6	58.0	31	5	US-09-574-376B-1240
17	11.6	58.0	37	5	US-09-708-690-18891
18	11.6	58.0	38	5	US-09-532-537B-1781
19	11.6	58.0	46	6	US-09-845-674-616
20	11.4	57.0	20	5	US-09-632-703-22
21	11.4	57.0	20	6	US-09-198-452A-1772
22	11.4	57.0	26	6	US-09-828-707-7
23	11.2	56.0	17	5	US-09-371-772B-5150
24	11.2	56.0	17	5	US-09-708-690-5150
25	11.2	56.0	19	7	US-09-747-391-16

26	11.2	56.0	20	6	US-09-198-452A-6351	Sequence 6351, Ap
27	11.2	56.0	25	6	US-09-881-565-19	Sequence 19, Appl
28	11.2	56.0	28	6	US-09-343-126B-50	Sequence 50, Appl
29	11.2	56.0	28	6	US-09-343-126B-51	Sequence 51, Appl
30	11.2	56.0	28	6	US-09-488-265-63	Sequence 63, Appl
31	11.2	56.0	31	6	US-09-801-274-736	Sequence 736, App
32	11.2	56.0	31	6	US-09-801-274-793	Sequence 793, App
33	11.2	56.0	31	6	US-09-801-274-1470	Sequence 1470, Ap
34	11.2	56.0	38	6	US-09-546-745A-5238	Sequence 5238, Ap
35	11.2	56.0	43	8	US-60-253-653-10876	Sequence 10876, A
36	11.2	56.0	50	5	US-09-699-011A-293	Sequence 293, App
37	11.2	56.0	50	6	US-09-881-565-22	Sequence 22, Appl
38	11.2	55.0	21	6	US-09-765-081-247	Sequence 247, App
39	11.2	55.0	23	7	US-09-851-271A-13	Sequence 13, Appl
40	11.2	55.0	28	7	US-09-735-271-1880	Sequence 1880, Ap
41	11.2	55.0	37	5	US-09-708-690-18282	Sequence 18282, A
42	11.2	55.0	37	5	US-09-708-690-18422	Sequence 18422, A
43	11.2	55.0	37	5	US-09-708-690-18674	Sequence 18674, A
44	11.2	55.0	37	5	US-09-708-690-18740	Sequence 18740, A
45	11.2	55.0	37	5	US-09-708-690-20466	Sequence 20466, A

ALIGNMENTS

RESULT 1

US-09-551-151-115

; Sequence 115, Application US/095511151

; GENERAL INFORMATION:

; APPLICANT: Borgford, Thor

; TITLE OF INVENTION: Ricin-Like Toxin Variants for Treatment of Cancer,

; FILE REFERENCE: 10447-011

; CURRENT APPLICATION NUMBER: US/09/551,151

; PRIOR FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: US 09/403,752

; PRIOR FILING DATE: 1999-10-29

; NUMBER OF SEQ ID NOS: 142

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 115

; LENGTH: 36

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Human Prostate-Specific Antigen linker region of pAP-290

US-09-551-151-115

Query Match 76.0%; Score 15.2; DB 7; Length 36;  
Best Local Similarity 85.0%; Pred. No. 4.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttctcccttcggtattctt 20

||||| ||||| ||||| ||

Db 14 ttctctcttcggtattttt 33

RESULT 2

US-09-726-774-70

; Sequence 70, Application US/09726774

; GENERAL INFORMATION:

; APPLICANT: Iversen, Patrick L.

; TITLE OF INVENTION: Antisense Antibacterial Method and

; FILE REFERENCE: 0450-0032.30

; CURRENT APPLICATION NUMBER: US/09/726,774

; CURRENT FILING DATE: 2000-11-29

; PRIOR APPLICATION NUMBER: US 60/168,150

; PRIOR FILING DATE: 1999-11-29

; NUMBER OF SEQ ID NOS: 139

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 70

; LENGTH: 21

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: antisense oligomer
US-09-726-774-70

Query Match      69.0%; Score 13.8; DB 6; Length 21;
Best Local Similarity 88.2%; Pred. No. 2.1e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttctccctccgtattc 18
   | ||||| |||||
Db 3 ttctccctccgtaattc 19

RESULT 3
US-09-900-590-83
; Sequence 83, Application US/09900590
; GENERAL INFORMATION:
; APPLICANT: Huse, William D.
; TITLE OF INVENTION: Anti-alpha V Beta 3 Recombinant Human
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/900,590
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/016,061
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-IX 2965
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..33
; SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-900-590-83

Query Match      69.0%; Score 13.8; DB 6; Length 33;
Best Local Similarity 88.2%; Pred. No. 2.2e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ttctccctccgtattc 20
   | | ||||| |||||
Db 3 TCTCATCCGTTATTCTT 19
```

```
RESULT 4
US-60-252-833-32245/c
; Sequence 32245, Application US/60252833
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R
; TITLE OF INVENTION: Compositions isolated from bovine
; FILE REFERENCE: 1052P2
; CURRENT APPLICATION NUMBER: US/60/252,833
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 43535
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32245
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Bovine
US-60-252-833-32245

Query Match      68.0%; Score 13.6; DB 8; Length 36;
Best Local Similarity 80.0%; Pred. No. 2.8e+03;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttctccctccgtattc 20
   ||||| |||||
Db 34 TTCTCCCTTGCTAGTCTT 15

RESULT 5
US-08-850-062A-75
; Sequence 75, Application US/08850062A
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; Goldgar, David E.
; Mikl, Yoshio
; Swenson, Jeff
; Kamb, Alexander
; Harshman, Keith D.
; Shattuck-Eidens, Donna M.
; Tavtigian, Sean V.
; Wiseman, Roger W.
; Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; Susceptibility Gene
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Manbeck, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,062A
; FILING DATE: 02-May-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,554
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
```

; APPLICATION NUMBER: US 08/289,221  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 2318-137  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-783-6040  
; TELEFAX: 202-9783-6031  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; SEQUENCE DESCRIPTION: SEQ ID NO: 75:  
US-08-850-062A-75

Query Match 63.0%; Score 12.6; DB 4; Length 30;  
Best Local Similarity 78.9%; Pred. No. 8.1e+03;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttctccctccggtattctt 20  
||||| ||||| | |||||  
DB 8 TCTCTCTTCCTCTCTCT 26

RESULT 6  
; Sequence 657, Application US/60278561  
; GENERAL INFORMATION:  
; APPLICANT: Morris, MacDonald  
; APPLICANT: Lal, Preeti  
; APPLICANT: Diep, Dinh  
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using  
; Nucleotide Sequence Databases, and Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms Identified Thereby  
; FILE REFERENCE: GX-0012-1 P  
; CURRENT APPLICATION NUMBER: US/60/278,561  
; CURRENT FILING DATE: 2001-03-23  
; NUMBER OF SEQ ID NOS: 15598  
; SOFTWARE: PERL Program  
; SEQ ID NO 657  
; LENGTH: 46  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: SNP00057293  
; NAME/KEY: snp  
; LOCATION: 21  
; OTHER INFORMATION: 017559.1, 21, G->A  
US-60-278-561-657

Query Match 63.0%; Score 12.6; DB 8; Length 46;  
Best Local Similarity 78.9%; Pred. No. 8.5e+03;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttctccctccggtattctt 19  
||||| ||||| | |||||  
DB 22 TTCCTCTTCCTCTCTCT 4

RESULT 7  
US-09-516-667-73/C  
; Sequence 73, Application US/09516667  
; GENERAL INFORMATION:

; APPLICANT: Inouye, Masayori  
; APPLICANT: Wang, Nan  
; APPLICANT: Yamanaka, Kunitoshi  
; TITLE OF INVENTION: COLD-SHOCK REGULATORY ELEMENTS, CONSTRUCTS THEREOF, AND  
; TITLE OF INVENTION: METHODS OF USE  
; FILE REFERENCE: 1053-00  
; CURRENT APPLICATION NUMBER: US/09/516,667  
; CURRENT FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 73  
; LENGTH: 47  
; TYPE: RNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: oligonucleotide  
US-09-516-667-73

Query Match 61.0%; Score 12.2; DB 6; Length 47;  
Best Local Similarity 82.4%; Pred. No. 1.3e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttctccctccggtattt 17  
||| ||||| | |||||  
DB 30 TTCCTCTTCCTCTCTT 14

RESULT 8  
US-09-516-667-35/C  
; Sequence 35, Application US/09516667  
; GENERAL INFORMATION:  
; APPLICANT: Inouye, Masayori  
; APPLICANT: Wang, Nan  
; APPLICANT: Yamanaka, Kunitoshi  
; TITLE OF INVENTION: COLD-SHOCK REGULATORY ELEMENTS, CONSTRUCTS THEREOF, AND  
; TITLE OF INVENTION: METHODS OF USE  
; FILE REFERENCE: 1053-00  
; CURRENT APPLICATION NUMBER: US/09/516,667  
; CURRENT FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 35  
; LENGTH: 48  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: oligonucleotide  
US-09-516-667-35

Query Match 61.0%; Score 12.2; DB 6; Length 48;  
Best Local Similarity 82.4%; Pred. No. 1.3e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttctccctccggtattt 17  
||| ||||| | |||||  
DB 29 TTCCTCTTCCTCTTATT 13

RESULT 9  
US-09-516-667-36  
; Sequence 36, Application US/09516667  
; GENERAL INFORMATION:  
; APPLICANT: Inouye, Masayori  
; APPLICANT: Wang, Nan  
; APPLICANT: Yamanaka, Kunitoshi  
; TITLE OF INVENTION: COLD-SHOCK REGULATORY ELEMENTS, CONSTRUCTS THEREOF, AND  
; TITLE OF INVENTION: METHODS OF USE  
; FILE REFERENCE: 1053-00  
; CURRENT APPLICATION NUMBER: US/09/516,667

; CURRENT FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 36  
; LENGTH: 48  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: oligonucleotide  
US-09-516-667-36

Query Match 61.0%; Score 12.2; DB 6; Length 48;  
Best Local Similarity 82.4%; Pred. No. 1.3e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttctcccttcggtatt 17  
||| ||||| |||||  
DB 24 ttccccccttcattatt 40

RESULT 10  
US-09-516-667-74/c  
; Sequence 74, Application US/09516667  
; GENERAL INFORMATION:  
; APPLICANT: Inouye, Masayori  
; APPLICANT: Wang, Nan  
; APPLICANT: Yamanaka, Kunitoshi  
; TITLE OF INVENTION: COLD-SHOCK REGULATORY ELEMENTS, CONSTRUCTS THEREOF, AND  
; TITLE OF INVENTION: METHODS OF USE  
; FILE REFERENCE: 1053-00  
; CURRENT APPLICATION NUMBER: US/09/516,667  
; CURRENT FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 74  
; LENGTH: 48  
; TYPE: RNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: oligonucleotide  
US-09-516-667-74

Query Match 61.0%; Score 12.2; DB 6; Length 48;  
Best Local Similarity 82.4%; Pred. No. 1.3e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttctcccttcggtatt 17  
||| ||||| |||||  
DB 30 TTCCCCCTTCATTATT 14

RESULT 11  
US-09-516-667-75/c  
; Sequence 75, Application US/09516667  
; GENERAL INFORMATION:  
; APPLICANT: Inouye, Masayori  
; APPLICANT: Wang, Nan  
; APPLICANT: Yamanaka, Kunitoshi  
; TITLE OF INVENTION: COLD-SHOCK REGULATORY ELEMENTS, CONSTRUCTS THEREOF, AND  
; TITLE OF INVENTION: METHODS OF USE  
; FILE REFERENCE: 1053-00  
; CURRENT APPLICATION NUMBER: US/09/516,667  
; CURRENT FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 75  
; LENGTH: 48  
; TYPE: RNA  
; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: oligonucleotide  
US-09-516-667-75

Query Match 61.0%; Score 12.2; DB 6; Length 48;  
Best Local Similarity 82.4%; Pred. No. 1.3e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttctcccttcggtatt 17  
||| ||||| |||||  
DB 30 TTCCCCCTTCATTATT 14

RESULT 12  
US-60-252-833-9824  
; Sequence 9824, Application US/60252833  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Grigor, Murray R  
; TITLE OF INVENTION: Compositions isolated from bovine  
; TITLE OF INVENTION: tissues and methods for their use.  
; FILE REFERENCE: 1052P2  
; CURRENT APPLICATION NUMBER: US/60/252,833  
; CURRENT FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 43535  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9824  
; LENGTH: 45  
; TYPE: DNA  
; ORGANISM: Bovine  
US-60-252-833-9824

Query Match 60.0%; Score 12; DB 8; Length 45;  
Best Local Similarity 75.0%; Pred. No. 1.6e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttctcccttcggtatttt 20  
||| ||||| |||||  
DB 3 tctgccttcgtgaagtctt 22

RESULT 13  
US-60-253-652-14008  
; Sequence 14008, Application US/60253652  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Grigor, Murray R  
; TITLE OF INVENTION: Compositions isolated from bovine  
; TITLE OF INVENTION: tissues and methods for their use.  
; FILE REFERENCE: 1055P2  
; CURRENT APPLICATION NUMBER: US/60/253,652  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 29954  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14008  
; LENGTH: 28  
; TYPE: DNA  
; ORGANISM: Bovine  
US-60-253-652-14008

Query Match 59.0%; Score 11.8; DB 8; Length 28;  
Best Local Similarity 86.7%; Pred. No. 1.9e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ctccttcggtatt 17  
| ||||| |||||  
DB 1 cacccttcggtctt 15

## RESULT 14

US-60-253-456-15489  
; Sequence 15489, Application US/60253456  
; GENERAL INFORMATION:  
; APPLICANT: Havukkala, Ilkka J,  
; TITLE OF INVENTION: Polynucleotides, isolated from  
; TITLE OF INVENTION: plants, and methods for their use.  
; FILE REFERENCE: 1054P1  
; CURRENT APPLICATION NUMBER: US/60/253,456  
; CURRENT FILING DATE: 2000-11-27  
; NUMBER OF SEQ ID NOS: 37096  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15489  
; LENGTH: 29  
; TYPE: DNA  
; ORGANISM: Pinus radiata  
US-60-253-456-15489

Query Match 58.0%; Score 11.6; DB 8; Length 29;  
Best Local Similarity 77.8%; Pred. No. 2.4e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tctcccttcggtattctt 19  
| | | | | | | | | |  
Db 1 tttccctacggtattct 18

## RESULT 15

US-09-574-376B-90/c  
; Sequence 90, Application US/09574376B  
; GENERAL INFORMATION:  
; APPLICANT: Warrington, Janet  
; APPLICANT: Shah, Nila  
; APPLICANT: Gingeras, Thomas Raymond  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Human Lymphoblast Polymorphisms  
; FILE REFERENCE: 3229.2  
; CURRENT APPLICATION NUMBER: US/09/574,376B  
; CURRENT FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 1330  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 90  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
US-09-574-376B-90

Query Match 58.0%; Score 11.6; DB 5; Length 31;  
Best Local Similarity 70.0%; Pred. No. 2.4e+04;  
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tctcccttcggtattctt 20  
| | | | | | | | | |  
Db 27 TTGTCACCTGCCTACTTCTT 8

Search completed: October 2, 2001, 16:55:00  
Job time: 17663 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 21:50:08 ; Search time 17695.9 Seconds  
(without alignments)  
16.701 Million cell updates/sec

Title: US-09-757-100B-6

Perfect score: 20

Sequence: 1 gaaactgcagaagcactga 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 738840595 residues

Total number of hits satisfying chosen parameters: 4519004

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_NA\_Main:\*

1: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq:\*

2: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq:\*

3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq:\*

4: /cgn2\_6/ptodata/1/pna/US08\_COMB.seq:\*

5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq:\*

6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq:\*

7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq:\*

8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq:\*

9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*

10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*

11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq:\*

12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq:\*

13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq:\*

14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq:\*

15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq:\*

16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq:\*

17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq:\*

18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq:\*

19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq:\*

20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq:\*

21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq:\*

22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq:\*

23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq:\*

24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq:\*

25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq:\*

26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq:\*

27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq:\*

28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq:\*

29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq:\*

30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq:\*

31: /cgn2\_6/ptodata/1/pna/US098\_COMB.seq:\*

32: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq:\*

33: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq:\*

34: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq:\*

35: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq:\*

36: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq:\*

37: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq:\*

38: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq:\*

39: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq:\*

40: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq:\*

41: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq:\*

42: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq:\*

43: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq:\*

44: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq:\*

45: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq:\*

46: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq:\*

47: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq:\*

48: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq:\*

49: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq:\*

50: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq:\*

51: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq:\*

52: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq:\*

53: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq:\*

54: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq:\*

55: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq:\*

56: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq:\*

57: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq:\*

58: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq:\*

59: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq:\*

60: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	20	1	PCT-US00-18999-6
2	20	100.0	20	17	US-09-377-310-6
3	20	100.0	20	29	US-09-757-100B-6
4	15.2	76.0	25	55	US-60-233-166-247462
5	15	75.0	15	1	PCT-US00-18999-26
6	15	75.0	15	17	US-09-377-310-26
7	15	75.0	15	29	US-09-757-100B-26
8	15	75.0	25	55	US-60-233-166-169359
9	15	75.0	25	55	US-60-233-166-169364
10	14.8	74.0	25	26	US-09-660-080-2372
11	14.8	74.0	25	26	US-09-660-220-7628
12	14.8	74.0	25	55	US-60-233-166-75968
13	14.4	72.0	25	55	US-60-233-166-247458
14	14.2	71.0	25	55	US-60-233-166-353412
15	14.2	71.0	25	55	US-60-233-357-7140
16	14.2	71.0	45	17	US-09-310-298-1603
17	13.8	69.0	20	31	US-09-844-521-16
18	13.8	69.0	25	55	US-60-233-166-250484
19	13.8	69.0	25	55	US-60-233-166-250515
20	13.8	69.0	25	55	US-60-233-620-33139
21	13.8	69.0	25	55	US-60-233-620-33145
22	13.8	69.0	25	55	US-60-233-620-33146
23	13.8	69.0	25	55	US-60-234-017-31881
24	13.8	69.0	25	55	US-60-234-017-31890
25	13.8	69.0	25	55	US-60-234-049-10183
26	13.6	68.0	25	55	US-60-233-166-142391
27	13.6	68.0	25	55	US-60-233-166-341643
28	13.6	68.0	25	55	US-60-233-166-358933
29	13.6	68.0	25	55	US-60-233-357-14127
30	13.6	68.0	25	55	US-60-234-017-150021
31	13.6	68.0	25	55	US-60-234-017-427143
32	13.6	68.0	25	55	US-60-234-017-427147
33	13.6	68.0	25	55	US-60-234-049-131450
34	13.6	68.0	42	17	US-09-318-138-15
35	13.4	67.0	19	28	US-09-703-708-12274
36	13.4	67.0	19	48	US-60-164-320-12274
37	13.4	67.0	19	50	US-60-183-791-12274
38	13.4	67.0	20	16	US-09-201-228A-4597
39	13.4	67.0	25	26	US-09-660-080-2373
40	13.4	67.0	25	55	US-60-233-166-75953
41	13.4	67.0	25	55	US-60-233-620-33138
42	13.4	67.0	36	16	US-09-232-278A-25
43	13.4	67.0	43	1	PCT-US96-13455-12
44	13.4	67.0	43	15	US-09-116-115-12
45	13.4	67.0	43	21	US-09-541-762-12

## ALIGNMENTS

RESULT 1  
PCT-US00-18999-6  
; Sequence 6, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase

; FILE REFERENCE: ISPH-0476  
; CURRENT FILING DATE: 2000-07-13  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-6

Query Match 100.0%; Score 20; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaactgcagaaggcactga 20  
Db 1 gaaactgcagaaggcactga 20  
|||||

RESULT 2  
US-09-377-310-6  
; Sequence 6, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-6

Query Match 100.0%; Score 20; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaactgcagaaggcactga 20  
Db 1 gaaactgcagaaggcactga 20  
|||||

RESULT 3  
US-09-757-100B-6  
; Sequence 6, Application US/09757100B

; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100B-6

Query Match 100.0%; Score 20; DB 29; Length 20;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaactgcagaaggcactga 20  
Db 1 gaaactgcagaaggcactga 20  
|||||

RESULT 4  
US-60-233-166-247462/C  
; Sequence 247462, Application US/60233166  
; GENERAL INFORMATION:  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat  
; FILE REFERENCE: 3112  
; CURRENT APPLICATION NUMBER: US/60/233,166  
; CURRENT FILING DATE: 2000-10-24  
; NUMBER OF SEQ ID NOS: 420907  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 247462  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AI043982  
US-60-233-166-247462

Query Match 76.0%; Score 15.2; DB 55; Length 25;  
Best Local Similarity 85.0%; Pred. No. 3.1e+03;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaactgcagaaggcactga 20  
Db 22 GAGTCAGCAGAAGGCACGCA 3  
|||

RESULT 5  
PCT-US00-18999-26  
; Sequence 26, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase

; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0476  
; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; CURRENT FILING DATE: 2000-07-13  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-26

Query Match 75.0%; Score 15; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.5e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aactgcagaagggcac 17  
|||||

Db 1 aactgcagaagggcac 15

## RESULT 6

US-09-377-310-26  
; Sequence 26, Application US/09377310A  
; GENERAL INFORMATION:

; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-26

Query Match 75.0%; Score 15; DB 17; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.5e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aactgcagaagggcac 17  
|||||

Db 1 aactgcagaagggcac 15

## RESULT 7

US-09-757-100B-26  
; Sequence 26, Application US/09757100B  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13

; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100B-26

Query Match 75.0%; Score 15; DB 29; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.5e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aactgcagaagggcac 17  
|||||

Db 1 aactgcagaagggcac 15

## RESULT 8

US-60-233-166-169359/c  
; Sequence 169359, Application US/60233166  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat  
; TITLE OF INVENTION: Genome  
; FILE REFERENCE: 3112  
; CURRENT APPLICATION NUMBER: US/60/233,166  
; CURRENT FILING DATE: 2000-10-24  
; NUMBER OF SEQ ID NOS: 420907  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 169359  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AA955408  
US-60-233-166-169359

Query Match 75.0%; Score 15; DB 55; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3.9e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaactgcagaagggc 15  
|||||

Db 18 GAAACTGCAGAAGGC 4

## RESULT 9

US-60-233-166-169364/c  
; Sequence 169364, Application US/60233166  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat  
; TITLE OF INVENTION: Genome  
; FILE REFERENCE: 3112  
; CURRENT APPLICATION NUMBER: US/60/233,166  
; CURRENT FILING DATE: 2000-10-24  
; NUMBER OF SEQ ID NOS: 420907  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 169364  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AA955408  
US-60-233-166-169364

Qy 3 aactgcagaaggcactga 20  
|||||  
Db 18 AACTGTAGAAGGCGCTGA 1

```
; CURRENT FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 420907
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 353412
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank U59126
US-60-233-166-353412
```

```
Query Match 71.0%; Score 14.2; DB 55; Length 25;
Best Local Similarity 84.2%; Pred. No. 9.7e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 2 aaactgcagaaggcactga 20
||||| ||||| ||| |
Db 2 aaactacagaaggtactca 20
```

```
RESULT 15
US-60-233-357-7140
; Sequence 7140, Application US/60233357
; GENERAL INFORMATION:
; APPLICANT: Affymetrix, Inc.
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat
; TITLE OF INVENTION: Neurobiology
; FILE REFERENCE: 3114
; CURRENT APPLICATION NUMBER: US/60/233,357
; CURRENT FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 21305
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7140
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank D49836
US-60-233-357-7140
```

```
Query Match 71.0%; Score 14.2; DB 55; Length 25;
Best Local Similarity 84.2%; Pred. No. 9.7e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 2 aaactgcagaaggcactga 20
||||| ||||| ||| |
Db 2 aaactacagaaggtactca 20
```

```
Search completed: October 2, 2001, 21:50:09
Job time: 24522 sec
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:54:59 ; Search time 2173.58 Seconds  
(without alignments)  
19.290 Million cell updates/sec

Title: US-09-757-100B-6

Perfect score: 20

Sequence: 1 gaaactgcagaagcactga 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1895189 seqs, 1048201267 residues

Total number of hits satisfying chosen parameters: 299022

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending\_Patents\_NA\_New.\*
- 1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq1.\*
  - 7: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq2.\*
  - 8: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	14.4	72.0	17	6	US-09-866-108-2022 Sequence 2022, Ap
c 2	14.4	72.0	17	6	US-09-866-108-2023 Sequence 2023, Ap
c 3	14.4	72.0	25	6	US-09-866-108-4951 Sequence 4951, Ap
c 4	14.4	72.0	25	6	US-09-866-108-4952 Sequence 4952, Ap
c 5	14.4	72.0	25	6	US-09-866-108-4953 Sequence 4953, Ap
c 6	14.4	72.0	25	6	US-09-866-108-4954 Sequence 4954, Ap
c 7	14.4	72.0	25	6	US-09-866-108-4955 Sequence 4955, Ap
c 8	14.4	72.0	25	6	US-09-866-108-4956 Sequence 4956, Ap
c 9	14.4	72.0	25	6	US-09-866-108-4957 Sequence 4957, Ap
c 10	14.4	72.0	25	6	US-09-866-108-4958 Sequence 4958, Ap
c 11	14.4	72.0	25	6	US-09-866-108-4959 Sequence 4959, Ap
c 12	14.4	72.0	25	6	US-09-866-108-4960 Sequence 4960, Ap
c 13	13.8	69.0	31	6	US-09-801-274-1595 Sequence 1595, Ap
c 14	13.4	67.0	17	6	US-09-866-108-2021 Sequence 2021, Ap
c 15	13.4	67.0	17	6	US-09-866-108-2024 Sequence 2024, Ap
c 16	13.4	67.0	25	6	US-09-866-108-4950 Sequence 4950, Ap
c 17	13.4	67.0	25	6	US-09-866-108-4961 Sequence 4961, Ap
c 18	13	65.0	17	6	US-09-866-108-2025 Sequence 2025, Ap
c 19	13	65.0	17	6	US-09-866-108-2026 Sequence 2026, Ap
c 20	13	65.0	25	6	US-09-866-108-4962 Sequence 4962, Ap
c 21	13	65.0	25	6	US-09-866-108-4963 Sequence 4963, Ap
c 22	12.6	63.0	24	7	US-09-919-478-18 Sequence 18, Appl
c 23	12.6	63.0	25	7	US-09-724-648-6 Sequence 6, Appl
c 24	12.6	63.0	26	6	US-09-902-214-52 Sequence 52, Appl
c 25	12.6	63.0	46	7	US-09-919-478-15 Sequence 15, Appl

ALIGNMENTS

RESULT 1

US-09-866-108-2022/c

; Sequence 2022, Application US/09866108

; GENERAL INFORMATION:

; APPLICANT: GU, Yizhong

; APPLICANT: JI, Yonggang

; APPLICANT: PENN, Sharron G.

; APPLICANT: HANZEL, David K.

; APPLICANT: RANK, David R.

; APPLICANT: CHEN, Wensheng

; TITLE OF INVENTION: SHANNON, Mark

; FILE REFERENCE: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE

; CURRENT APPLICATION NUMBER: AEOICA-7

; CURRENT FILING DATE: 2001-05-25

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 60/266,860

; PRIOR FILING DATE: 2001-02-05

; NUMBER OF SEQ ID NOS: 15752

; SOFTWARE: Aeomica Sequence Listing Engine

; SEQ ID NO 2022

Sequence 2020, Ap  
Sequence 4949, Ap  
Sequence 7, Appli  
Sequence 33, Appl  
Sequence 223, App  
Sequence 3, Appli  
Sequence 2, Appli  
Sequence 2027, Ap  
Sequence 6556, Ap  
Sequence 4964, Ap  
Sequence 45, Appl  
Sequence 1777, Ap  
Sequence 370, App  
Sequence 370, App  
Sequence 11980, A  
Sequence 14257, A  
Sequence 18382, A  
Sequence 2, Appli  
Sequence 39, Appli

; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-2022

Query Match 72.0%; Score 14.4; DB 6; Length 17;  
Best Local Similarity 93.8%; Pred. No. 1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ctgcagaaggcactga 20  
|||||  
Db 17 CTGCAGAGGCACCGA 2

## RESULT 2

US-09-866-108-2023/c  
; Sequence 2023, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.  
; APPLICANT: HANZEL, David K.  
; APPLICANT: RANK, David R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: SHANNON, Mark  
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
; FILE REFERENCE: AEOICA-7  
; CURRENT APPLICATION NUMBER: US/09/866,108  
; CURRENT FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 60/266,860  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aecomica Sequence Listing Engine  
; SEQ ID NO 2023  
; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-2023

Query Match 72.0%; Score 14.4; DB 6; Length 17;  
Best Local Similarity 93.8%; Pred. No. 1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ctgcagaaggcactga 20  
|||||  
Db 16 CTGCAGAGGCACCGA 1

## RESULT 3

US-09-866-108-4951/c  
; Sequence 4951, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong  
; APPLICANT: JI, Yonggang  
; APPLICANT: PENN, Sharron G.  
; APPLICANT: HANZEL, David K.  
; APPLICANT: RANK, David R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: SHANNON, Mark  
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
; FILE REFERENCE: AEOICA-7  
; CURRENT APPLICATION NUMBER: US/09/866,108  
; CURRENT FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 60/266,860  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 15752  
; SOFTWARE: Aecomica Sequence Listing Engine  
; SEQ ID NO 4951  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-866-108-4951

Query Match 72.0%; Score 14.4; DB 6; Length 25;  
Best Local Similarity 93.8%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ctgcagaaggcactga 20  
|||||  
Db 25 CTGCAGAGGCACCGA 10

## RESULT 4

US-09-866-108-4952/c  
; Sequence 4952, Application US/09866108  
; GENERAL INFORMATION:  
; APPLICANT: GU, Yizhong



```

> PRIOR APPLICATION NUMBER: US 60/207,456
> PRIOR FILING DATE: 2000-05-26
> PRIOR APPLICATION NUMBER: GB 24263.6
> PRIOR FILING DATE: 2000-10-04
> PRIOR APPLICATION NUMBER: US 60/236,359
> PRIOR FILING DATE: 2000-09-27
> PRIOR APPLICATION NUMBER: PCT/US01/006666
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006667
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006664
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006659
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006665
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006668
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006663
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006662
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006661
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006670
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: US 60/234,687
> PRIOR FILING DATE: 2000-09-21
> PRIOR APPLICATION NUMBER: US 60/266,860
> PRIOR FILING DATE: 2001-02-05
> NUMBER OF SEQ ID NOS: 15752
> SOFTWARE: Aeomica Sequence Listing Engine
> SEQ ID NO 4953
> LENGTH: 25
> TYPE: DNA
> ORGANISM: Homo sapiens
US-09-866-108-4953

Query Match 72.0%; Score 14.4
Best Local Similarity 93.8%; Pred. No. 1
Matches 15; Conservative 0; Mismatch

Qy 5 ctgcagaaggcactga 20
Db 23 CTGCAGAGGCCCGA 8
|||||
|||||

RESULT 6
US-09-866-108-4954/c
> Sequence 4954, Application US/09866108
> GENERAL INFORMATION:
> APPLICANT: GU, Yizhong
> APPLICANT: Ji, Yonggang
> APPLICANT: PENN, Sharon G.
> APPLICANT: HANZEL, David K.
> APPLICANT: RANK, David R.
> APPLICANT: CHEN, Wensheng
> APPLICANT: SHANNON, Mark
> TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSION
> FILE REFERENCE: AEOMICA-7
> CURRENT APPLICATION NUMBER: US/09/866,108
> CURRENT FILING DATE: 2001-05-25
> PRIOR APPLICATION NUMBER: US 60/207,456
> PRIOR FILING DATE: 2000-05-26
> PRIOR APPLICATION NUMBER: GB 24263.6
> PRIOR FILING DATE: 2000-10-04
> PRIOR APPLICATION NUMBER: US 60/236,359
> PRIOR FILING DATE: 2000-09-27
> PRIOR APPLICATION NUMBER: PCT/US01/006666
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/006667
> PRIOR FILING DATE: 2001-01-30

```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aecomica Sequence Listing Engine
; SEQ ID NO 4954
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-4954
```

```
Query Match 72.0%; Score 14.4; DB 6; Length 25;
Best Local Similarity 93.8%; Pred. No. 1.1e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 5 ctgcagaagcactga 20
||| ||||| ||||| ||
Db 22 CTGCAGAGGCACCGA 7
```

```
RESULT 7
US-09-866-108-4955/c
; Sequence 4955, Application US/09866108
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aecomica Sequence Listing Engine
; SEQ ID NO 4955
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-4955
```

```
Query Match 72.0%; Score 14.4; DB 6; Length 25;
Best Local Similarity 93.8%; Pred. No. 1.1e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 5 ctgcagaagcactga 20
||| ||||| ||||| ||
Db 21 CTGCAGAGGCACCGA 6
```

```
RESULT 8
US-09-866-108-4956/c
; Sequence 4956, Application US/09866108
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
```



Sequence 4959, Application US/09866108  
GENERAL INFORMATION:  
APPLICANT: GU, Yizhong  
APPLICANT: JI, Yonggang  
APPLICANT: PENN, Sharon G.  
APPLICANT: HANZEL, David K.  
APPLICANT: RANK, David R.  
APPLICANT: CHEN, Wensheng  
APPLICANT: SHANNON, Mark  
TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE  
FILE REFERENCE: AEOMICA-7  
CURRENT APPLICATION NUMBER: US/09/866,108  
CURRENT FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 60/266,860  
PRIOR FILING DATE: 2001-02-05  
NUMBER OF SEQ ID NOS: 15752  
SOFTWARE: Aecomica Sequence Listing Engine  
SEQ ID NO 4959  
LENGTH: 25  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-866-108-4959

Query Match 72.0%; Score 14.4; DB 6; Length 25;  
Best Local Similarity 93.8%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ctgcagaaggcactga 20  
|||||

Db 17 CTGCAGAGGCACCGA 2

RESULT 12  
US-09-866-108-4960/c  
Sequence 4960, Application US/09866108  
GENERAL INFORMATION:  
APPLICANT: GU, Yizhong  
APPLICANT: JI, Yonggang  
APPLICANT: PENN, Sharon G.  
APPLICANT: HANZEL, David K.  
APPLICANT: RANK, David R.  
APPLICANT: CHEN, Wensheng  
APPLICANT: SHANNON, Mark  
TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE

FILE REFERENCE: AEOMICA-7  
CURRENT APPLICATION NUMBER: US/09/866,108  
CURRENT FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 60/266,860  
PRIOR FILING DATE: 2001-02-05  
NUMBER OF SEQ ID NOS: 15752  
SOFTWARE: Aecomica Sequence Listing Engine  
SEQ ID NO 4960  
LENGTH: 25  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-866-108-4960

Query Match 72.0%; Score 14.4; DB 6; Length 25;  
Best Local Similarity 93.8%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ctgcagaaggcactga 20  
|||||

Db 16 CTGCAGAGGCACCGA 1

RESULT 13  
US-09-801-274-1595  
Sequence 1595, Application US/09801274  
GENERAL INFORMATION:  
APPLICANT: Cargill, Michele  
APPLICANT: Ireland, James S.  
APPLICANT: Lander, Eric S.  
TITLE OF INVENTION: HUMAN SINGLE NUCLEOTIDE POLYMORPHISMS  
FILE REFERENCE: 2825.2009-001  
CURRENT APPLICATION NUMBER: US/09/801,274  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: US 60/187,510  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 60/206,129  
PRIOR FILING DATE: 2000-05-22  
NUMBER OF SEQ ID NOS: 1802  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1595  
LENGTH: 31  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-801-274-1595

Query Match 69.0%; Score 13.8; DB 6; Length 31;  
Best Local Similarity 78.9%; Pred. No. 2.3e+03;  
Matches 15; Conservative 1; Mismatches 3; Indels

**Qy**      1 gaaactgcagaaggcactg 19  
          || ||||| |  
**Dd**      2 gacactgcaggagggmgtg 20

RESULT 14

US-09-866-108-2021/c

Sequence 2021, Application US/09866108

GENERAL INFORMATION:

APPLICANT: GU, Yizhong

APPLICANT: JI, Yonggang

APPLICANT: PENN, Sharron G.

APPLICANT: HANZEL, David K.

APPLICANT: RANK, David R.

APPLICANT: CHEN, Wensheng

APPLICANT: SHANNON, Mark

TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE

FILE REFERENCE: AEMICA-7

CURRENT APPLICATION NUMBER: US/09/866,108

CURRENT FILING DATE: 2001-05-25

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/006666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 60/266,860

PRIOR FILING DATE: 2001-02-05

NUMBER OF SEQ ID NOS: 15752

SOFTWARE: Aemica Sequence Listing Engine

SEQ ID NO 2021

LENGTH: 17

TYPE: DNA

ORGANISM: Homo sapiens

US-09-866-108-2021

Query Match 67.0%; Score 13.4; DB 6;  
Best Local Similarity 93.3%; Pred. No. 3.2e+03;  
Matches 14: Conservative 0; Mismatches 1; Indels

Qy	6	tgcagaaggcactga	20
Db	17	TGCAGAAGGCACCGA	3

```

RESULT      15
US-09-866-108-2024/c
; Sequence 2024, Application US/09866108
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYO5N-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOmica-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeo mica Sequence Listing Engine
; SEQ ID NO 2024
; LENGTH: 17
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-2024

```

```

Query Match      67.0%; Score 13.4; DB 6; Length 17;
Best Local Similarity 93.3%; Pred. No. 3.2e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      5  ctgcagaaggcactg 19
          |||||
Db      15  CTGCAGAAAGGCACCG 1

```

Search completed: October 2, 2001, 16:55:00  
Job time: 17663 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:54:58 ; Search time 2173.58 Seconds  
(without alignments)  
19.290 Million cell updates/sec

Title: US-09-757-100B-4

Perfect score: 20

Sequence: 1 ggcgcgtgaagcgaaggca 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1895189 seqs, 1048201267 residues

Total number of hits satisfying chosen parameters: 299022

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_NA\_New.\*

- 1: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 8: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12.2	61.0	43	8	US-60-252-833-26301
2	12	60.0	35	7	US-09-475-947A-204
3	12	60.0	50	6	US-09-442-849B-1086
4	12	60.0	50	7	US-09-443-199C-1086
C 5	11.8	59.0	17	6	US-09-817-879-1279
C 6	11.8	59.0	17	6	US-09-817-879-1504
C 7	11.8	59.0	17	6	US-09-817-879-1505
8	11.8	59.0	17	6	US-09-817-879-3051
9	11.8	59.0	17	6	US-09-817-879-3276
10	11.8	59.0	31	5	US-09-574-376B-157
C 11	11.8	59.0	31	7	US-09-724-671-4282
C 12	11.6	58.0	22	7	US-09-918-889-26
C 13	11.6	58.0	22	7	US-09-919-042-26
C 14	11.6	58.0	24	5	US-09-929-404-282
C 15	11.6	58.0	41	1	PCT-US01-25861-1
C 16	11.6	58.0	41	5	US-09-932-581-1
C 17	11.4	57.0	48	6	US-09-477-962-23
C 18	11.4	57.0	42	8	US-60-252-833-6500
C 19	11.2	56.0	20	6	US-09-800-629A-185
C 20	11.2	56.0	43	8	US-60-253-653-10876
C 21	11.2	56.0	40	6	US-60-253-378-27275
C 22	11	55.0	26	6	US-09-198-452A-6446
23	11	55.0	26	7	US-09-145-916B-22
24	11	55.0	31	5	US-09-574-376B-1105
25	11	55.0	31	6	US-09-817-879-6300

26	11	55.0	35	7	US-09-864-785-3865	Sequence 3865, Ap
C 27	11	55.0	37	6	US-09-595-639-6	Sequence 6, Appli
C 28	11	55.0	38	5	US-09-371-772B-12058	Sequence 12058, A
C 29	11	55.0	38	5	US-09-708-690-14335	Sequence 14335, A
30	11	55.0	41	8	US-60-253-653-23884	Sequence 23884, A
31	11	55.0	43	7	US-09-880-505-81	Sequence 81, Appl
32	11	55.0	45	7	US-09-633-739-10	Sequence 10, Appl
C 33	11	55.0	45	8	US-60-253-654-19893	Sequence 19893, A
C 34	11	55.0	45	8	US-60-255-592-19893	Sequence 19893, A
C 35	10.8	54.0	15	5	US-09-274-553C-355	Sequence 355, App
C 36	10.8	54.0	15	5	US-09-274-553C-1025	Sequence 1025, Ap
C 37	10.8	54.0	16	5	US-09-765-400-33	Sequence 33, Appl
C 38	10.8	54.0	17	6	US-09-817-879-1280	Sequence 1280, Ap
39	10.8	54.0	17	6	US-09-817-879-3050	Sequence 3050, Ap
40	10.8	54.0	17	6	US-09-817-879-3277	Sequence 3277, Ap
C 41	10.8	54.0	18	6	US-09-863-806-180	Sequence 180, App
C 42	10.8	54.0	18	6	US-09-863-806-186	Sequence 186, App
C 43	10.8	54.0	23	6	US-09-803-110-1474	Sequence 1474, Ap
C 44	10.8	54.0	32	1	PCT-US01-1337A-40	Sequence 40, Appl
45	10.8	54.0	36	7	US-09-846-040-165	Sequence 165, App

ALIGNMENTS

RESULT 1

US-60-252-833-26301  
; Sequence 26301, Application US/60252833  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Grigor, Murray R  
; TITLE OF INVENTION: Compositions isolated from bovine  
; TITLE OF INVENTION: tissues and methods for their use.  
; FILE REFERENCE: 1052P2  
; CURRENT APPLICATION NUMBER: US/60/252,833  
; CURRENT FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 43535  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 26301  
; LENGTH: 43  
; TYPE: DNA  
; ORGANISM: Bovine  
US-60-252-833-26301

Query Match 61.0%; Score 12.2; DB 8; Length 43;  
Best Local Similarity 82.4%; Pred. No. 9.1e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggcgcgtgaagcgaag 17  
| | | | | | | | | | | | | | | | | | |  
Db 1 ggcgcgtgaagcgaag 17

RESULT 2

US-09-475-947A-204  
; Sequence 204, Application US/09475947A  
; GENERAL INFORMATION:  
; APPLICANT: Garner, Harold R.  
; APPLICANT: Wren, Jonathan D.  
; APPLICANT: Minna, John D.  
; TITLE OF INVENTION: Polymorphic Repeats in Human Genes  
; FILE REFERENCE: UTSD0667  
; CURRENT APPLICATION NUMBER: US/09/475,947A  
; CURRENT FILING DATE: 1999-12-31  
; NUMBER OF SEQ ID NOS: 346  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 204  
; LENGTH: 35  
; TYPE: DNA  
; ORGANISM: human  
US-09-475-947A-204

Query Match 60.0%; Score 12; DB 7; Length 35;  
Best Local Similarity 75.0%; Pred. No. 1.1e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggccgcgtgaagcggaagca 20  
||| |||| |||| |||| ||||  
Db 8 ggccgcgtgaagcggaagca 27

## RESULT 3

US-09-442-849B-1086

; Sequence 1086, Application US/09442849B  
; GENERAL INFORMATION:  
; APPLICANT: Shinketsu, Richard A.  
; APPLICANT: Leach, Martin  
; TITLE OF INVENTION: Nucleic Acids Containing Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms and Methods of Use Thereof  
; FILE REFERENCE: 15966-534C  
; CURRENT APPLICATION NUMBER: US/09/442.849B  
; PRIOR FILING DATE: 1999-11-17  
; PRIOR APPLICATION NUMBER: 60/109,024  
; PRIOR FILING DATE: 1998-11-17  
; PRIOR APPLICATION NUMBER: 09/443,199  
; PRIOR FILING DATE: 1999-11-16  
; PRIOR APPLICATION NUMBER: 09/442, 129  
; PRIOR FILING DATE: 1999-11-16  
; NUMBER OF SEQ ID NOS: 1592  
; SOFTWARE: Curagen Patent Formatter Version 0.9  
; SEQ ID NO 1086  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (26)...(0)  
; OTHER INFORMATION: 2 of 2 allelic variants (1085 is other entry)  
; NAME/KEY: misc\_feature  
; LOCATION: (25)...(26)  
; OTHER INFORMATION: nucleotide deleted between bases 25 and 26  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Accession number cg44923068  
US-09-442-849B-1086

Query Match 60.0%; Score 12; DB 6; Length 50;  
Best Local Similarity 75.0%; Pred. No. 1.2e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggccgcgtgaagcggaagca 20  
||| |||| |||| |||| ||||  
Db 23 ggactcgtgaagcggaagca 42

## RESULT 4

US-09-443-199C-1086

; Sequence 1086, Application US/09443199C  
; GENERAL INFORMATION:  
; APPLICANT: Shinketsu, Richard A.  
; APPLICANT: Leach, Martin  
; TITLE OF INVENTION: Nucleic Acids Containing Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms and Methods of Use Thereof  
; FILE REFERENCE: 15966-534A  
; CURRENT APPLICATION NUMBER: US/09/443.199C  
; PRIOR FILING DATE: 1999-11-16  
; PRIOR APPLICATION NUMBER: 60/109,024  
; PRIOR FILING DATE: 1998-11-17  
; NUMBER OF SEQ ID NOS: 1272  
; SOFTWARE: Curagen Patent Formatter Version 0.9  
; SEQ ID NO 1086  
; LENGTH: 50  
; TYPE: DNA

; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (26)...(0)  
; OTHER INFORMATION: 2 of 2 allelic variants (1085 is other entry)  
; NAME/KEY: misc\_feature  
; LOCATION: (25)...(26)  
; OTHER INFORMATION: nucleotide deleted between bases 25 and 26  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Accession number cg44923068  
US-09-443-199C-1086

Query Match 60.0%; Score 12; DB 7; Length 50;  
Best Local Similarity 75.0%; Pred. No. 1.2e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggccgcgtgaagcggaagca 20  
||| |||| |||| |||| ||||  
Db 23 ggactcgtgaagcggaagca 42

## RESULT 5

US-09-817-879-1279/c  
; Sequence 1279, Application US/09817879  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals Inc.  
; TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Rel  
; TITLE OF INVENTION: Hepatitis C Virus Infection  
; FILE REFERENCE: MBH00-801-F  
; CURRENT APPLICATION NUMBER: US/09/817.879  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 9703  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1279  
; LENGTH: 17  
; TYPE: RNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: oligonucleotide substrate  
US-09-817-879-1279

Query Match 59.0%; Score 11.8; DB 6; Length 17;  
Best Local Similarity 86.7%; Pred. No. 1.4e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 ccgtgaagcggaagcc 19  
||| |||| |||| ||||  
Db 17 CCGGTAGCGTAGGC 3

## RESULT 6

US-09-817-879-1504/c  
; Sequence 1504, Application US/09817879  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals Inc.  
; TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Rel  
; TITLE OF INVENTION: Hepatitis C Virus Infection  
; FILE REFERENCE: MBH00-801-F  
; CURRENT APPLICATION NUMBER: US/09/817.879  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 9703  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1504  
; LENGTH: 17  
; TYPE: RNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; NAME/KEY: misc\_feature



LOCATION:  
OTHER INFORMATION: oligonucleotide substrate  
US-09-817-879-1504

Query Match 59.0%; Score 11.8; DB 6; Length 17;  
Best Local Similarity 86.7%; Pred. No. 1.4e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ccgtgaagcgaagc 19  
||| ||||| |||

DB 17 CCGGAAGCGAAGC 3

## RESULT 7

US-09-817-879-1505/c  
Sequence 1505, Application US/09817879  
GENERAL INFORMATION:  
APPLICANT: Ribozyme Pharmaceuticals Inc.  
TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Related to Hepatitis C Virus Infection  
FILE REFERENCE: MH800-801-F  
CURRENT APPLICATION NUMBER: US/09/817,879  
CURRENT FILING DATE: 2001-03-26  
NUMBER OF SEQ ID NOS: 9703  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1505  
LENGTH: 17  
TYPE: RNA  
ORGANISM: artificial sequence  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION:  
OTHER INFORMATION: oligonucleotide substrate  
US-09-817-879-1505

Query Match 59.0%; Score 11.8; DB 6; Length 17;  
Best Local Similarity 86.7%; Pred. No. 1.4e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gcgcctgaagcgaa 16  
||| ||||| |||

DB 15 GCCCGGAGCGAA 1

## RESULT 8

US-09-817-879-3051  
Sequence 3051, Application US/09817879  
GENERAL INFORMATION:  
APPLICANT: Ribozyme Pharmaceuticals Inc.  
TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Related to Hepatitis C Virus Infection  
FILE REFERENCE: MH800-801-F  
CURRENT APPLICATION NUMBER: US/09/817,879  
CURRENT FILING DATE: 2001-03-26  
NUMBER OF SEQ ID NOS: 9703  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3051  
LENGTH: 17  
TYPE: RNA  
ORGANISM: artificial sequence  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION:  
OTHER INFORMATION: oligonucleotide substrate  
US-09-817-879-3051

Query Match 59.0%; Score 11.8; DB 6; Length 17;  
Best Local Similarity 86.7%; Pred. No. 1.4e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ccgtgaagcgaagc 19  
||| ||||| |||

DB 2 ccgcgaagcgaagc 16

## RESULT 9

US-09-817-879-3276  
Sequence 3276, Application US/09817879  
GENERAL INFORMATION:  
APPLICANT: Ribozyme Pharmaceuticals Inc.  
TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Related to Hepatitis C Virus Infection  
FILE REFERENCE: MH800-801-F  
CURRENT APPLICATION NUMBER: US/09/817,879  
CURRENT FILING DATE: 2001-03-26  
NUMBER OF SEQ ID NOS: 9703  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3276  
LENGTH: 17  
TYPE: RNA  
ORGANISM: artificial sequence  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION:  
OTHER INFORMATION: oligonucleotide substrate  
US-09-817-879-3276

Query Match 59.0%; Score 11.8; DB 6; Length 17;  
Best Local Similarity 80.0%; Pred. No. 1.4e+04;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 ccgtgaagcgaagc 19  
||| ||||| |||

DB 2 ccgugagcguaagc 16

## RESULT 10

US-09-574-376B-157  
Sequence 157, Application US/09574376B  
GENERAL INFORMATION:  
APPLICANT: Warrington, Janet  
APPLICANT: Shah, Nila  
APPLICANT: Gingeras, Thomas Raymond  
APPLICANT: Affymetrix, Inc.  
TITLE OF INVENTION: Human Lymphoblast Polymorphisms  
FILE REFERENCE: 3229.2  
CURRENT APPLICATION NUMBER: US/09/574,376B  
CURRENT FILING DATE: 2000-05-19  
NUMBER OF SEQ ID NOS: 1330  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 157  
LENGTH: 31  
TYPE: DNA  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION:  
US-09-574-376B-157

Query Match 59.0%; Score 11.8; DB 5; Length 31;  
Best Local Similarity 86.7%; Pred. No. 1.4e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 cgtgaagcgaagcga 20  
||| ||||| |||||

DB 17 cgagaagagaagcga 31

## RESULT 11

US-09-724-671-4282/c  
Sequence 4282, Application US/09724671  
GENERAL INFORMATION:  
APPLICANT: Watson, James D

Query Match  
58.0%; Score 11.6; DB 5; Length 24;

Best Local Similarity 77.8%; Pred. No. 1.8e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaagg 18  
| ||| ||||| |||  
Db 6 gtcgctgtgaagccaacg 23

RESULT 15  
PCT-US01-25861-1/c  
; Sequence 1. Application PC/TUS0125861  
; GENERAL INFORMATION:  
; APPLICANT: Sierra Sciences, Inc.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING  
; FILE REFERENCE: TELOMERASE REVERSE TRANSCRIPTASE (TERT) EXPRESSION  
; CURRENT APPLICATION NUMBER: PCT/US01/25861  
; CURRENT FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/227,865  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 60/230,174  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/238,345  
; PRIOR FILING DATE: 2000-10-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 41  
; TYPE: DNA  
; ORGANISM: human  
PCT-US01-25861-1

Query Match 58.0%; Score 11.6; DB 1; Length 41;  
Best Local Similarity 77.8%; Pred. No. 1.8e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ggcgcgtgaagcgaaggc 19  
| |||| | | | | |  
Db 24 GCCTCGCGGAGGAGAGGC 7

Search completed: October 2, 2001, 16:54:59  
Job time: 17662 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 21:50:08 ; Search time 17695.9 Seconds  
(without alignments)  
16.701 Million cell updates/sec

Title: US-09-757-100B-4

Perfect score: 20

Sequence: 1 ggcgcgctgaagcgaagca 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 4519004

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_NA\_Main:\*

1: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq:\*

2: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq:\*

3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq:\*

4: /cgn2\_6/ptodata/1/pna/US08\_COMB.seq:\*

5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq:\*

6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq:\*

7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq:\*

8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq:\*

9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*

10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*

11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq:\*

12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq:\*

13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq:\*

14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq:\*

15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq:\*

16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq:\*

17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq:\*

18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq:\*

19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq:\*

20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq:\*

21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq:\*

22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq:\*

23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq:\*

24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq:\*

25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq:\*

26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq:\*

27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq:\*

28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq:\*

29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq:\*

30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq:\*

31: /cgn2\_6/ptodata/1/pna/US098\_COMB.seq:\*

32: /cgn2\_6/ptodata/1/pna/US099\_COMB.seq:\*

33: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq:\*

34: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq:\*

35: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq:\*

36: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq:\*

37: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq:\*

38: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq:\*

39: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq:\*

40: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq:\*

41: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq:\*

42: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq:\*

43: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq:\*

44: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq:\*

45: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq:\*

46: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq:\*

47: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq:\*

48: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq:\*

49: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq:\*

50: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq:\*

51: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq:\*

52: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq:\*

53: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq:\*

54: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq:\*

55: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq:\*

56: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq:\*

57: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq:\*

58: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq:\*

59: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq:\*

60: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	20	1	PCT-US00-18999-4
2	20	100.0	20	17	US-09-377-310-4
3	20	100.0	20	29	US-09-757-100B-4
4	15	75.0	15	1	PCT-US00-18999-24
5	15	75.0	15	17	US-09-377-310-4
6	15	75.0	15	29	US-09-757-100B-24
7	14.8	74.0	45	8	US-08-448-223-13
8	14.8	74.0	45	8	US-08-448-223-14
9	14.4	72.0	25	55	US-60-234-017-94926
10	13.6	68.0	25	55	US-60-234-017-952075
11	13.6	68.0	30	31	US-09-812-913-11
12	13.6	68.0	41	18	US-09-404-520-40172
13	13.6	68.0	41	18	US-09-404-520-40466
14	13.2	66.0	25	55	US-60-232-638-45792
15	13.2	66.0	25	55	US-60-233-166-104969
16	13.2	66.0	25	55	US-60-233-166-104976
17	13.2	66.0	25	55	US-60-233-166-104980
18	13.2	66.0	25	55	US-60-233-166-105010
19	13	65.0	25	55	US-60-234-017-568725
20	13	65.0	25	55	US-60-234-017-568728
21	13	65.0	25	55	US-60-234-017-568731
22	13	65.0	25	55	US-60-234-017-568734
23	12.8	64.0	25	26	US-09-660-080-19599
24	12.8	64.0	25	26	US-09-660-220-12063
25	12.8	64.0	25	55	US-60-233-166-200509
26	12.8	64.0	25	55	US-60-233-166-309711
27	12.8	64.0	25	55	US-60-233-620-2588
28	12.8	64.0	25	55	US-60-233-620-2593
29	12.8	64.0	25	55	US-60-233-620-78332
30	12.8	64.0	25	55	US-60-233-620-83352
31	12.8	64.0	25	55	US-60-234-049-4741
32	12.8	64.0	25	55	US-60-234-049-4743
33	12.8	64.0	25	55	US-60-234-049-4744
34	12.8	64.0	25	55	US-60-234-049-4746
35	12.8	64.0	25	55	US-60-234-049-4754
36	12.8	64.0	25	55	US-60-234-049-4788
37	12.8	64.0	25	55	US-60-234-049-4789
38	12.8	64.0	25	55	US-60-234-049-4790
39	12.8	64.0	25	55	US-60-234-049-4791
40	12.8	64.0	25	55	US-60-234-049-4792
41	12.8	64.0	25	55	US-60-234-049-4793
42	12.8	64.0	25	55	US-60-234-049-4795
43	12.8	64.0	25	55	US-60-234-049-5481
44	12.8	64.0	25	55	US-60-234-049-59410
45	12.8	64.0	25	55	US-60-234-049-73099

## ALIGNMENTS

RESULT 1  
PCT-US00-18999-4  
; Sequence 4, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0476  
; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-4

Query Match 100.0%; Score 20; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcgcgtgaagcgaagca 20  
| | | | | | | | | | | | | | | | | |  
Db 1 ggcgcgtgaagcgaagca 20

RESULT 2  
US-09-377-310-4  
; Sequence 4, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-4

Query Match 100.0%; Score 20; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcgcgtgaagcgaagca 20  
| | | | | | | | | | | | | | | | | |  
Db 1 ggcgcgtgaagcgaagca 20

RESULT 3  
US-09-757-100B-4  
; Sequence 4, Application US/09757100B

; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100B-4

Query Match 100.0%; Score 20; DB 29; Length 20;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcgcgtgaagcgaagca 20  
| | | | | | | | | | | | | | | | | |  
Db 1 ggcgcgtgaagcgaagca 20

RESULT 4  
PCT-US00-18999-24  
; Sequence 24, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0476  
; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; CURRENT FILING DATE: 2000-07-13  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-24

Query Match 75.0%; Score 15; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cgcgcgtgaagcgaag 17  
| | | | | | | | | | | | | | | | | |  
Db 1 cgcgcgtgaagcgaag 15

RESULT 5  
US-09-377-310-24  
; Sequence 24, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.

APPLICANT: Gaarde, William A.  
TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
FILE REFERENCE: ISPH-0389  
CURRENT APPLICATION NUMBER: US/09/377,310A  
CURRENT FILING DATE: 1999-08-19  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 24  
LENGTH: 15  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: antisense sequence  
US-09-377-310-24

Query Match 75.0%; Score 15; DB 17; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cgcgcgtgaagcgaag 17  
|||||

DB 1 cgcgcgtgaagcgaag 15

## RESULT 6

US-09-757-100B-24  
Sequence 24, Application US/09757100B  
GENERAL INFORMATION:  
APPLICANT: Monia, Brett P.  
APPLICANT: Gaarde, William A.  
APPLICANT: Nero, Pamela S.  
TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
TITLE OF INVENTION: Expression  
FILE REFERENCE: ISPH-0533  
CURRENT APPLICATION NUMBER: US/09/757,100B  
CURRENT FILING DATE: 2001-03-15  
PRIOR APPLICATION NUMBER: 09/377,310  
PRIOR FILING DATE: 1999-08-19  
PRIOR APPLICATION NUMBER: PCT/US00/18999  
PRIOR FILING DATE: 2000-07-13  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 24  
LENGTH: 15  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: antisense sequence  
US-09-757-100B-24

Query Match 75.0%; Score 15; DB 29; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cgcgcgtgaagcgaag 17  
|||||

DB 1 cgcgcgtgaagcgaag 15

## RESULT 7

US-08-448-223-13  
Sequence 13, Application US/08448223  
GENERAL INFORMATION:  
APPLICANT: Abramson, Richard D.  
APPLICANT: Gelfand, David H.  
APPLICANT: Kalman, Lisa V.  
APPLICANT: Reichert, Fred L.  
TITLE OF INVENTION: Thermostable DNA Polymerases With  
TITLE OF INVENTION: Altered Discrimination Properties  
NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,223  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Sias, Stacey R.  
REGISTRATION NUMBER: 32,630  
REFERENCE/DOCKET NUMBER: 9134  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 814-2863  
TELEFAX: (510) 814-2977  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-448-223-13

Query Match 74.0%; Score 14.8; DB 8; Length 45;  
Best Local Similarity 88.9%; Pred. No. 4.9e+03;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gcgcgcgtgaagcgaagc 19  
|||||

DB 3 gcgcgcgtgaagcgaagc 20

## RESULT 8

US-08-448-223-14  
Sequence 14, Application US/08448223  
GENERAL INFORMATION:  
APPLICANT: Abramson, Richard D.  
APPLICANT: Gelfand, David H.  
APPLICANT: Kalman, Lisa V.  
APPLICANT: Reichert, Fred L.  
TITLE OF INVENTION: Thermostable DNA Polymerases With  
TITLE OF INVENTION: Altered Discrimination Properties  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,223  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Sias, Stacey R.  
REGISTRATION NUMBER: 32,630  
REFERENCE/DOCKET NUMBER: 9134

```

:
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (510) 814-2863
: TELEFAX: (510) 814-2977
:
: INFORMATION FOR SEQ ID NO: 14:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 45 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: US-08-448-223-14

```

```
Query Match          74.0%; Score 14.8; DB 8; Length 45;
Best Local Similarity 88.9%; Pred. NO. 4.9e+03;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Oy 2 gcgccgtgaagcgaagc 19  
 ||||| |||||  
 Db 3 GCGCCGCCAAGCGAAGC 20

RESULT 9  
US-60-234-017-94926/c

```

% REPLICATION 05/002354017
%
% GENERAL INFORMATION:
% APPLICANT: Mitmann, M
% APPLICANT: Affymetrix, Inc.
% TITLE OF INVENTION: Methods of Genetic Analysis of Mus
% TITLE OF INVENTION: musculus
% FILE REFERENCE: 3115
% CURRENT APPLICATION NUMBER: US/60/234,017
% CURRENT FILING DATE: 2000-09-20
% NUMBER OF SEQ IDS: 605887
% SOFTWARE: FastSeq for Windows Version 4.0
% SEQ ID NO 94926

```

```
Query Match          72.0%; Score 14.4; DB 55; Length 25;
Best Local Similarity 93.8%; Pred. No. 7.6e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 4 gccgtgaagcgaaggc 19  
||||| |||||  
Db 18 GCCGTGCAGCGAAGGC 3

```

RESULT 10
US-60-234-017-352075/c
; Sequence 352075, Application US/60234017
; GENERAL INFORMATION:
; APPLICANT: Mittmann, M
; APPLICANT: Affymetrix, Inc.
; TITLE OF INVENTION: Methods of Genetic Analysis of Mus
; TITLE OF INVENTION: musculus
; FILE REFERENCE: 3115
; CURRENT APPLICATION NUMBER: US/60/234,017
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 605887
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 352075

```

Query Match 68.0%; Score 13.6; DB 55; Length 25;  
Best Local Similarity 80.0%; Pred. No. 1.9e+04;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggcgccgtgaagcgaaggca 20  
| | | | | | | | | |  
Db 21 GCCTCAGTGACGCCGAAGGCA 2

```

RESULT 11
US-09-812-913-11/c
; Sequence 11, Application US/09812913
; GENERAL INFORMATION:
; APPLICANT: Weeks, Donald
; APPLICANT: Cerutti, Heriberto
; TITLE OF INVENTION: Acetolactate synthase gene and construct for the transformation
; TITLE OF INVENTION: extranuclear genomes
; FILE REFERENCE: N1231-210
; CURRENT APPLICATION NUMBER: US/09/812,913
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,668
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
; US-09-812-913-11

```

Query Match 68.0%; Score 13.6; DB 31; Length 30;  
Best Local Similarity 80.0%; Pred. No. 1.9e+04;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgccgtgaagcgaaggca 20  
|||  
Db 24 GCGCCGTGCAGCGCGCGCA 5

```

RESULT 12
US-09-404-520-40172
; Sequence 40172, Application US/09404520
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Ghodssi, Azita
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: McIninch, James
; APPLICANT: Timberlake, William E.
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: Emmericella nidulans Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15498)A
; CURRENT APPLICATION NUMBER: US/09/404,520
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 44345
; SEQ ID NO 40172
; LENGTH: 41
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-09-404-520-40172

```

Query Match 68.0%; Score 13.6; DB 18; Length 41;  
Best Local Similarity 80.0%; Pred. No. 1.9e+04;  
Matches 16: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggcgcgtgaagcgaaggca 20  
| | | | | | | | | |  
pb 13 gccgccatgagccaaaggca 32



```
RESULT 13
US-09-404-520-40466
; Sequence 40466, Application US/09404520
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Ghodssi, Azita
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: McIninch, James
; APPLICANT: Timberlake, William E.
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: Emericella nidulans Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15498)A
; CURRENT APPLICATION NUMBER: US/09/404,520
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 44345
; SEQ ID NO 40466
; LENGTH: 41
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-09-404-520-40466

Query Match      68.0%; Score 13.6; DB 18; Length 41;
Best Local Similarity 80.0%; Pred. No. 1.9e+04;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcgcgtgaagcgaaaggca 20
   | |||| |||| | ||||
Db 13 ggcgcctgaagaggaggagca 32

RESULT 14
US-60-232-638-45792
; Sequence 45792, Application US/60232638
; GENERAL INFORMATION:
; APPLICANT: Mittmann
; TITLE OF INVENTION: Methods of Genetic Analysis of Yeast
; FILE REFERENCE: 3110
; CURRENT APPLICATION NUMBER: US/60/232,638
; CURRENT FILING DATE: 2000-09-14
; NUMBER OF SEQ ID NOS: 138410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45792
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: SGD YDL200C
US-60-232-638-45792

Query Match      66.0%; Score 13.2; DB 55; Length 25;
Best Local Similarity 83.3%; Pred. No. 2.9e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgcccgtaagcgaaaggca 20
   | |||| | |||| |
Db 5 cgccgtcaaccgaaaggga 22

RESULT 15
US-60-233-166-104969/c
; Sequence 104969, Application US/60233166
; GENERAL INFORMATION:
; APPLICANT: Mittmann
; APPLICANT: Affymetrix, Inc.
; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat
; FILE REFERENCE: 3112
; CURRENT APPLICATION NUMBER: US/60/233,166
; CURRENT FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 420907
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 104969
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank AA859343
US-60-233-166-104969
```

```
Query Match      66.0%; Score 13.2; DB 55; Length 25;
Best Local Similarity 83.3%; Pred. No. 2.9e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ggcgcgtgaagcgaaaggc 19
   | |||| | |||| |
Db 24 GGGCCGCTAGCGAAGGC 7
```

Search completed: October 2, 2001, 21:50:08  
Job time: 24521 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 15:01:27 ; Search time 17695.9 Seconds  
(without alignments)  
16.701 Million cell updates/sec

Title: US-09-757-100B-3  
Perfect score: 20  
Sequence: 1 ccgcgggtcacagtggctcg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1415048 seqs, 7388405095 residues  
Total number of hits satisfying chosen parameters: 4519004

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Pending\_Patents\_NA\_Main:\*

1: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq:\*

2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq:\*

3: /cgn2\_6/ptodata/1/pna/US08\_COMB.seq:\*

4: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq:\*

5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq:\*

6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq:\*

7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq:\*

8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq:\*

9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*

10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*

11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq:\*

12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq:\*

13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq:\*

14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq:\*

15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq:\*

16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq:\*

17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq:\*

18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq:\*

19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq:\*

20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq:\*

21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq:\*

22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq:\*

23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq:\*

24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq:\*

25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq:\*

26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq:\*

27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq:\*

28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq:\*

29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq:\*

30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq:\*

31: /cgn2\_6/ptodata/1/pna/US098\_COMB.seq:\*

32: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq:\*

33: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq:\*

34: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq:\*

35: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq:\*

36: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq:\*

37: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq:\*

38: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq:\*

39: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq:\*

40: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq:\*

41: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq:\*

42: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq:\*

43: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq:\*

44: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq:\*

45: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq:\*

46: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq:\*

47: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq:\*

48: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq:\*

49: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq:\*

50: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq:\*

51: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq:\*

52: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq:\*

53: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq:\*

54: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq:\*

55: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq:\*

56: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq:\*

57: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq:\*

58: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq:\*

59: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq:\*

60: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	20	100.0	20	1	PCT-US00-18999-3	Sequence 3, Appli
2	20	100.0	20	17	US-09-377-310-3	Sequence 3, Appli
3	20	100.0	20	29	US-09-757-100B-3	Sequence 3, Appli
4	15	75.0	15	1	PCT-US00-18999-23	Sequence 23, Appli
5	15	75.0	15	17	US-09-377-310-23	Sequence 23, Appli
6	15	75.0	15	29	US-09-757-100B-23	Sequence 23, Appli
7	14.8	74.0	25	26	US-09-660-220-37376	Sequence 37376, A
8	14.8	74.0	25	26	US-09-660-220-37377	Sequence 37377, A
9	14.8	74.0	25	55	US-60-233-166-185790	Sequence 185790, A
10	14.2	71.0	25	55	US-60-234-017-55657	Sequence 55657, A
11	14.2	71.0	25	55	US-60-233-166-279728	Sequence 279728, A
12	14	70.0	25	55	US-60-233-166-319513	Sequence 319513, A
13	13.8	69.0	25	55	US-60-233-166-10846	Sequence 10846, A
14	13.8	69.0	25	55	US-60-234-017-248844	Sequence 248844, A
15	13.6	68.0	40	9	US-08-594-866-2	Sequence 2, Appli
16	13.6	68.0	25	55	US-60-233-166-418139	Sequence 418139, A
17	13.6	68.0	25	55	US-09-724-866-18962	Sequence 18962, A
18	13.6	68.0	35	29	US-09-724-866A-18962	Sequence 18962, A
19	13.6	68.0	35	29	US-60-171-432-18962	Sequence 18962, A
20	13.6	68.0	35	49	US-09-083-194A-6	Sequence 6, Appli
21	13.6	68.0	39	14	US-09-326-203A-31	Sequence 31, Appli
22	13.4	67.0	25	55	US-60-232-638-136549	Sequence 136549, A
23	13.4	67.0	25	55	US-60-233-166-296555	Sequence 296555, A
24	13.4	67.0	25	55	US-60-234-017-217856	Sequence 217856, A
25	13.4	67.0	25	55	US-60-233-166-410983	Sequence 410983, A
26	13.2	66.0	25	55	US-60-233-166-413618	Sequence 413618, A
27	13.2	66.0	25	55	US-60-234-017-125344	Sequence 125344, A
28	13.2	66.0	25	55	US-60-234-017-242830	Sequence 242830, A
29	13.2	66.0	29	17	US-09-304-232-546	Sequence 546, App
30	13.2	66.0	31	14	US-09-006-298-11	Sequence 11, Appli
31	13.2	66.0	39	18	US-09-405-459-173	Sequence 173, App
32	13.2	66.0	50	21	US-09-549-848B-57	Sequence 57, Appli
33	13.2	66.0	50	21	US-09-660-220-37375	Sequence 37375, A
34	12.8	64.0	25	55	US-60-233-166-30520	Sequence 30520, A
35	12.8	64.0	25	55	US-60-233-166-40877	Sequence 40877, A
36	12.8	64.0	25	55	US-60-233-166-185404	Sequence 185404, A
37	12.8	64.0	25	55	US-60-233-166-235736	Sequence 235736, A
38	12.8	64.0	25	55	US-60-233-166-272127	Sequence 272127, A
39	12.8	64.0	25	55	US-60-233-166-272785	Sequence 272785, A
40	12.8	64.0	25	55	US-60-233-166-341785	Sequence 341785, A
41	12.8	64.0	25	55	US-60-233-166-341789	Sequence 341789, A
42	12.8	64.0	25	55	US-60-233-166-404414	Sequence 404414, A
43	12.8	64.0	25	55	US-60-234-017-480869	Sequence 480869, A

## ALIGNMENTS

RESULT 1  
PCT-US00-18999-3  
; Sequence 3, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0476  
; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-3

Query Match 100.0%; Score 20; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtgtcg 20  
Db 1 ccgcgggctcacagtgtcg 20

RESULT 2  
US-09-377-310-3  
; Sequence 3, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-3

Query Match 100.0%; Score 20; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtgtcg 20  
Db 1 ccgcgggctcacagtgtcg 20

RESULT 3  
US-09-757-100B-3  
; Sequence 3, Application US/09757100B

; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100B-3

Query Match 100.0%; Score 20; DB 29; Length 20;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtgtcg 20  
Db 1 ccgcgggctcacagtgtcg 20

RESULT 4  
PCT-US00-18999-23  
; Sequence 23, Application PC/TUS0018999  
; GENERAL INFORMATION:  
; APPLICANT: Isis Pharmaceuticals, Inc.  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0476  
; CURRENT APPLICATION NUMBER: PCT/US00/18999  
; CURRENT FILING DATE: 2000-07-13  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
PCT-US00-18999-23

Query Match 75.0%; Score 15; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 4.1e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gcggggctcacagtgg 17  
Db 1 gcggggctcacagtgg 15

RESULT 5  
US-09-377-310-23  
; Sequence 23, Application US/09377310A  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.

; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310A  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-23

Query Match 75.0%; Score 15; DB 17; Length 15;  
Best Local Similarity 100.0%; Pred. No. 4.1e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcggggctcacagtgg 17  
| | | | | | | | | | | | | | |  
DB 1 gcggggctcacagtgg 15

RESULT 6  
US-09-757-100b-23  
; Sequence 23, Application US/09757100B  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; APPLICANT: Nero, Pamela S.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0533  
; CURRENT APPLICATION NUMBER: US/09/757,100B  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 09/377,310  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/US00/18999  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-757-100b-23

Query Match 75.0%; Score 15; DB 29; Length 15;  
Best Local Similarity 100.0%; Pred. No. 4.1e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcggggctcacagtgg 17  
| | | | | | | | | | | | | | |  
DB 1 gcggggctcacagtgg 15

RESULT 7  
US-09-660-220-37376  
; Sequence 37376, Application US/09660220  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann et al.  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of Human  
; FILE REFERENCE: 3102.1  
; CURRENT APPLICATION NUMBER: US/09/660,220  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 60/164,973

; PRIOR FILING DATE: 1999-11-11  
; NUMBER OF SEQ ID NOS: 140981  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 37376  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank L11931  
US-09-660-220-37376

Query Match 74.0%; Score 14.8; DB 26; Length 25;  
Best Local Similarity 88.9%; Pred. No. 5.1e+03;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ccgggggtcacagtgtgc 19  
| | | | | | | | | | | | | | |  
DB 6 ccgggggtcacagtgtgc 23

RESULT 8  
US-09-660-220-37377  
; Sequence 37377, Application US/09660220  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann et al.  
; TITLE OF INVENTION: Methods of Genetic Analysis of Human  
; FILE REFERENCE: 3102.1  
; CURRENT APPLICATION NUMBER: US/09/660,220  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 60/164,973  
; PRIOR FILING DATE: 1999-11-11  
; NUMBER OF SEQ ID NOS: 140981  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 37377  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank L11931  
US-09-660-220-37377

Query Match 74.0%; Score 14.8; DB 26; Length 25;  
Best Local Similarity 88.9%; Pred. No. 5.1e+03;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ccgggggtcacagtgtgc 19  
| | | | | | | | | | | | | | |  
DB 2 ccgggggtcacagtgtgc 19

RESULT 9  
US-60-233-166-185790  
; Sequence 185790, Application US/60233166  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann  
; APPLICANT: Affymetrix, Inc.  
; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat  
; FILE REFERENCE: 3112  
; CURRENT APPLICATION NUMBER: US/60/233,166  
; CURRENT FILING DATE: 2000-10-24  
; NUMBER OF SEQ ID NOS: 420907  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 185790  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AA963963  
US-60-233-166-185790

Query Match 74.0%; Score 14.8; DB 55; Length 25;  
Best Local Similarity 88.9%; Pred. No. 5.1e+03;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 gcggggtcacagtgtgcg 20  
||| |||| |||| ||||  
Db 2 gcggactcacgtgtgcg 19

RESULT 10  
US-60-234-017-55657  
; Sequence 55657, Application US/60234017  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann, M  
; TITLE OF INVENTION: Methods of Genetic Analysis of Mus  
; FILE REFERENCE: 3115  
; CURRENT APPLICATION NUMBER: US/60/234,017  
; CURRENT FILING DATE: 2000-09-20  
; NUMBER OF SEQ ID NOS: 605887  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 55657  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank U76759  
US-60-234-017-55657

Query Match 71.0%; Score 14.2; DB 55; Length 25;  
Best Local Similarity 84.2%; Pred. No. 1e+04;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cgcgggtcacagtgtgcg 20  
|| |||| |||| ||||  
Db 2 cgtgggtcacagagtgcg 20

RESULT 11  
US-60-234-017-75527/c  
; Sequence 75527, Application US/60234017  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann, M  
; TITLE OF INVENTION: Methods of Genetic Analysis of Mus  
; FILE REFERENCE: 3115  
; CURRENT APPLICATION NUMBER: US/60/234,017  
; CURRENT FILING DATE: 2000-09-20  
; NUMBER OF SEQ ID NOS: 605887  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 75527  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AF048838  
US-60-234-017-75527

Query Match 71.0%; Score 14.2; DB 55; Length 25;  
Best Local Similarity 84.2%; Pred. No. 1e+04;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcgggtcacagtgtgcg 19  
||| |||| |||| ||||  
Db 19 CCGCGGACCCACAGTGGAC 1

RESULT 12  
US-60-233-166-279728/c  
; Sequence 279728, Application US/60233166  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann  
; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat  
; FILE REFERENCE: 3112  
; CURRENT APPLICATION NUMBER: US/60/233,166  
; CURRENT FILING DATE: 2000-10-24  
; NUMBER OF SEQ ID NOS: 420907  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 279728  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AF639107  
US-60-233-166-279728

Query Match 70.0%; Score 14; DB 55; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.2e+04;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ggcacacagtgtgc 19  
||||| |||| ||||  
Db 21 GGCTCACAGTGTGC 8

RESULT 13  
US-60-233-166-319513  
; Sequence 319513, Application US/60233166  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann  
; TITLE OF INVENTION: Methods of Genetic Analysis of the Rat  
; FILE REFERENCE: 3112  
; CURRENT APPLICATION NUMBER: US/60/233,166  
; CURRENT FILING DATE: 2000-10-24  
; NUMBER OF SEQ ID NOS: 420907  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 319513  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank L02896  
US-60-233-166-319513

Query Match 69.0%; Score 13.8; DB 55; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.6e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 cggggtcacagtgtgcg 20  
||| |||| |||| ||||  
Db 2 cggactcacagtgtgcg 18

RESULT 14  
US-60-234-017-10846  
; Sequence 10846, Application US/60234017  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann, M  
; TITLE OF INVENTION: Methods of Genetic Analysis of Mus  
; FILE REFERENCE: 3115  
; CURRENT APPLICATION NUMBER: US/60/234,017  
; CURRENT FILING DATE: 2000-09-20

; NUMBER OF SEQ ID NOS: 605887  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10846  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AF055638  
US-60-234-017-10846

Query Match 69.0%; Score 13.8; DB 55; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.6e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cgcgggctcacagtgg 18  
||| |||||  
Db 6 cgcgatctcacagtgg 22

RESULT 15  
US-60-234-017-248844  
; Sequence 248844, Application US/60234017  
; GENERAL INFORMATION:  
; APPLICANT: Mittmann, M  
; TITLE OF INVENTION: Methods of Genetic Analysis of Mus  
; TITLE OF INVENTION: musculus  
; FILE REFERENCE: 3115  
; CURRENT APPLICATION NUMBER: US/60/234,017  
; CURRENT FILING DATE: 2000-09-20  
; NUMBER OF SEQ ID NOS: 605887  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 248844  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank AW049900  
US-60-234-017-248844

Query Match 69.0%; Score 13.8; DB 55; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.6e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 cgggctcacagtggcg 20  
||| |||||  
Db 7 cggggtctcagtggcg 23

Search completed: October 2, 2001, 21:50:08  
Job time: 24521 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 12:00:37 ; Search time 2173.58 Seconds  
(without alignments)  
19.290 Million cell updates/sec

Title: US-09-757-100B-3

Perfect score: 20

Sequence: 1 ccgcgggtcacagtgtcg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1895189 seqs, 1048201267 residues

Total number of hits satisfying chosen parameters: 299022

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_New.\*

- 1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq1.\*
- 7: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq2.\*
- 8: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	13.2	66.0	50	7	US-09-688-069-57
2	13.2	66.0	50	7	US-09-688-071-57
3	12.2	61.0	41	7	US-09-688-069-51
4	12.2	61.0	41	7	US-09-688-071-51
5	12.2	61.0	45	7	US-09-688-069-53
6	12.2	61.0	45	7	US-09-688-071-53
7	11.6	58.0	25	7	US-09-735-271-2037
8	11.6	58.0	25	6	US-09-879-341-12
9	11.6	58.0	26	6	US-09-915-060-9
10	11.6	58.0	30	7	US-09-915-060-43
11	11.6	58.0	30	8	US-09-298-340-21
12	11.6	58.0	30	7	US-09-787-252-17
13	11.4	57.0	18	7	US-09-801-274-629
14	11.4	57.0	31	6	US-09-735-271-1558
15	11.4	57.0	40	7	US-09-336-643-50
16	11.4	57.0	45	6	US-09-371-772B-4196
17	11.2	56.0	17	5	US-09-708-690-4196
18	11.2	56.0	17	5	US-09-825-497-23
19	11.2	56.0	20	6	US-09-883-152-55
20	11.2	56.0	22	6	US-09-801-274-1447
21	11.2	56.0	31	6	US-09-688-069-55
22	11.2	56.0	38	7	US-09-688-069-59
23	11.2	56.0	38	7	US-09-688-071-55
24	11.2	56.0	38	7	US-09-688-071-55
25	11.2	56.0	38	7	US-09-688-071-59

c	26	11.2	56.0	40	5	US-09-941-992-324	Sequence 324, App
	27	11.2	56.0	47	1	PCT-US01-23545-17	Sequence 17, Appl
	28	11.2	56.0	47	1	PCT-US01-23707-17	Sequence 17, Appl
	29	11.2	56.0	47	7	US-09-912-436-17	Sequence 17, Appl
c	30	11.2	56.0	48	6	US-09-803-110-7471	Sequence 7471, Ap
c	31	11.2	56.0	48	7	US-09-724-750-7833	Sequence 7833, Ap
	32	11	55.0	20	6	US-09-851-896-76	Sequence 76, Appl
c	33	11	55.0	22	7	US-09-748-537-5	Sequence 5, Appli
c	34	11	55.0	37	6	US-09-803-110-1481	Sequence 1481, Ap
c	35	11	55.0	47	8	US-60-253-457-20631	Sequence 20631, A
c	36	11	55.0	49	8	US-60-252-833-32799	Sequence 32799, A
	37	11	55.0	50	6	US-09-504-576A-10254	Sequence 10254, A
c	38	10.8	54.0	16	5	US-09-371-772B-5913	Sequence 5913, Ap
c	39	10.8	54.0	16	5	US-09-708-690-5913	Sequence 5913, Ap
c	40	10.8	54.0	17	5	US-09-371-772B-5002	Sequence 5002, Ap
c	41	10.8	54.0	17	5	US-09-708-690-5002	Sequence 5002, Ap
c	42	10.8	54.0	17	5	US-09-708-690-7832	Sequence 7832, Ap
c	43	10.8	54.0	17	6	US-09-546-745A-2312	Sequence 2312, Ap
c	44	10.8	54.0	17	6	US-09-817-879-143	Sequence 143, App
c	45	10.8	54.0	17	6	US-09-817-879-4412	Sequence 4412, Ap

ALIGNMENTS

RESULT 1  
US-09-688-069-57  
; Sequence 57, Application US/09688069  
; GENERAL INFORMATION:  
; APPLICANT: Subramaniam, S.; Slater, S.; Karberg, K.; Chen, R.; Valentin, H.; Wong,  
; TITLE OF INVENTION: Nucleic Acid Sequences to Proteins Involved in Tocopherol Synt  
; FILE REFERENCE: 16515.054  
; CURRENT APPLICATION NUMBER: US/09/688,069  
; CURRENT FILING DATE: 2000-10-14  
; NUMBER OF SEQ ID NOS: 114  
; SEQ ID NO 57  
; LENGTH: 50  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Oligonucleotides  
US-09-688-069-57

Query Match 66.0% Score 13.2; DB 7; Length 50;  
Best Local Similarity 83.3%; Pred. No. 2.9e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcgggtcacagtgt 18  
||||| ||||| |||||  
Db 5 ccgcgggtcacagtgt 22

RESULT 2  
US-09-688-071-57  
; Sequence 57, Application US/09688071  
; GENERAL INFORMATION:  
; APPLICANT: Lassner, M.; Post-Beittenmiller, M.; Savidge, B.; Weiss, J.; Mitsky, T.  
; TITLE OF INVENTION: Nucleic Acid Sequences to Proteins Involved in Tocopherol Synt  
; FILE REFERENCE: 16515.055  
; CURRENT APPLICATION NUMBER: US/09/688,071  
; CURRENT FILING DATE: 2000-10-14  
; PRIOR APPLICATION NUMBER: US 60/129,899  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: US 60/146,461  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: US 09/549,848  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: WO PCT/US00/10368  
; PRIOR FILING DATE: 2000-09-14  
; NUMBER OF SEQ ID NOS: 114  
; SEQ ID NO 57

; LENGTH: 50  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Oligonucleotides  
US-09-688-071-57

Query Match 66.0%; Score 13.2; DB 7; Length 50;

Best Local Similarity 83.3%; Pred. No. 2.9e+03;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcggggtcacagtgtt 18

||||| | |||| | ||

Db 5 ccgcggccgcacaatggt 22

RESULT 3

US-09-688-069-51

; Sequence 51, Application US/09688069

; GENERAL INFORMATION:

; APPLICANT: Subramaniam, S.; Slater, S.; Karberg, K.; Chen, R.; Valentin, H.; Wong, Y.

; TITLE OF INVENTION: Nucleic Acid Sequences to Proteins Involved in Tocopherol Synthes

; FILE REFERENCE: 16515.054

; CURRENT APPLICATION NUMBER: US/09/688.069

; CURRENT FILING DATE: 2000-10-14

; NUMBER OF SEQ ID NOS: 114

; SEQ ID NO 51

; LENGTH: 41

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic Oligonucleotides

US-09-688-069-51

Query Match 61.0%; Score 12.2; DB 7; Length 41;

Best Local Similarity 82.4%; Pred. No. 9.4e+03;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcggggtcacagtgg 17

||||| | |||| | ||

Db 5 ccgcggccgcacaatgg 21

RESULT 4

US-09-688-071-51

; Sequence 51, Application US/09688071

; GENERAL INFORMATION:

; APPLICANT: Lassarner, M.; Post-Beittenmiller, M.; Savidge, B.; Weiss, J.; Mitsky, T.;

; APPLICANT: Valentin, H.

; TITLE OF INVENTION: Nucleic Acid Sequences to Proteins Involved in Tocopherol Synthes

; FILE REFERENCE: 16515.055

; CURRENT APPLICATION NUMBER: US/09/688.071

; CURRENT FILING DATE: 2000-10-14

; PRIOR APPLICATION NUMBER: US 60/129,899

; PRIOR FILING DATE: 1999-04-15

; PRIOR APPLICATION NUMBER: US 60/146,461

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: US 09/549,848

; PRIOR FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: WO PCT/US00/10368

; PRIOR FILING DATE: 2000-09-14

; NUMBER OF SEQ ID NOS: 114

; SEQ ID NO 51

; LENGTH: 41

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic Oligonucleotides

US-09-688-071-51

Query Match 61.0%; Score 12.2; DB 7; Length 41;

Best Local Similarity 82.4%; Pred. No. 9.4e+03;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcggggtcacagtgg 17

||||| | |||| | ||

Db 5 ccgcggccgcacaatgg 21

RESULT 5

US-09-688-069-53

; Sequence 53, Application US/09688069

; GENERAL INFORMATION:

; APPLICANT: Subramaniam, S.; Slater, S.; Karberg, K.; Chen, R.; Valentin, H.; Wong,

; TITLE OF INVENTION: Nucleic Acid Sequences to Proteins Involved in Tocopherol Synt

; FILE REFERENCE: 16515.054

; CURRENT APPLICATION NUMBER: US/09/688.069

; CURRENT FILING DATE: 2000-10-14

; NUMBER OF SEQ ID NOS: 114

; SEQ ID NO 53

; LENGTH: 45

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic Oligonucleotides

US-09-688-069-53

Query Match 61.0%; Score 12.2; DB 7; Length 45;

Best Local Similarity 82.4%; Pred. No. 9.4e+03;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcggggtcacagtgg 17

||||| | |||| | ||

Db 5 ccgcggccgcacaatgg 21

RESULT 6

US-09-688-071-53

; Sequence 53, Application US/09688071

; GENERAL INFORMATION:

; APPLICANT: Lassarner, M.; Post-Beittenmiller, M.; Savidge, B.; Weiss, J.; Mitsky, T.

; APPLICANT: Valentin, H.

; TITLE OF INVENTION: Nucleic Acid Sequences to Proteins Involved in Tocopherol Synt

; FILE REFERENCE: 16515.055

; CURRENT APPLICATION NUMBER: US/09/688.071

; CURRENT FILING DATE: 2000-10-14

; PRIOR APPLICATION NUMBER: US 60/129,899

; PRIOR FILING DATE: 1999-04-15

; PRIOR APPLICATION NUMBER: US 60/146,461

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: US 09/549,848

; PRIOR FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: WO PCT/US00/10368

; PRIOR FILING DATE: 2000-09-14

; NUMBER OF SEQ ID NOS: 114

; SEQ ID NO 53

; LENGTH: 45

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic Oligonucleotides

US-09-688-071-53

Query Match 61.0%; Score 12.2; DB 7; Length 45;

Best Local Similarity 82.4%; Pred. No. 9.4e+03;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgcggggtcacagtgg 17

||||| | |||| | ||

Db 5 ccgcggccgcacaatgg 21

## RESULT 7

US-09-735-271-2037  
; Sequence 2037, Application US/09735271  
; GENERAL INFORMATION:  
; APPLICANT: Daly, Mark J.  
; APPLICANT: Hudson, Thomas J.  
; APPLICANT: Lander, Eric S.  
; APPLICANT: Rioux, John  
; APPLICANT: Siminovich, Kathy  
; TITLE OF INVENTION: IBD-RELATED POLYMORPHISMS  
; FILE REFERENCE: 2825.1025-002  
; CURRENT APPLICATION NUMBER: US/09/735,271  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 60/170,257  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 60/196,046  
; PRIOR FILING DATE: 2000-04-10  
; NUMBER OF SEQ ID NOS: 2058  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2037  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-735-271-2037

Query Match 59.0%; Score 11.8; DB 7; Length 25;

Best Local Similarity 86.7%; Pred. No. 1.5e+04; Indels 0; Gaps 0;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgcgggtcacagtg 16

I I I I I I I I I I I

Db 2 ctcgggtcacagtg 16

## RESULT 8

US-09-879-341-12  
; Sequence 12, Application US/09879341  
; GENERAL INFORMATION:  
; APPLICANT: Kster, Hubert  
; TITLE OF INVENTION: DNA Diagnostics Based on Mass Spectrometry  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe LLP  
; STREET: 4350 La Jolla Village Drive  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122-1246  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/09/879,341  
; FILING DATE: 11-Jun-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/786,416  
; FILING DATE: 28-FEB-2001  
; APPLICATION NUMBER: 09/287,679  
; FILING DATE: 06-APR-1999  
; APPLICATION NUMBER: 08/617,256  
; FILING DATE: 18-MAR-1996  
; APPLICATION NUMBER: 08/406,199  
; FILING DATE: 03-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24736-2002M  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-879-341-12

Query Match 58.0%; Score 11.6; DB 6; Length 25;

Best Local Similarity 77.8%; Pred. No. 1.8e+04; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 gcggggtcacagtggtcg 20

I I I I I I I I I I I I

Db 6 GCGGCAACACAGTCGTG 23

## RESULT 9

US-09-879-341-10/c  
; Sequence 10, Application US/09879341  
; GENERAL INFORMATION:  
; APPLICANT: Kster, Hubert  
; TITLE OF INVENTION: DNA Diagnostics Based on Mass Spectrometry  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe LLP  
; STREET: 4350 La Jolla Village Drive  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122-1246  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/09/879,341  
; FILING DATE: 11-Jun-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/786,416  
; FILING DATE: 28-FEB-2001  
; APPLICATION NUMBER: 09/287,679  
; FILING DATE: 06-APR-1999  
; APPLICATION NUMBER: 08/617,256  
; FILING DATE: 18-MAR-1996  
; APPLICATION NUMBER: 08/406,199  
; FILING DATE: 03-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24736-2002M  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-879-341-10

Query Match 58.0%; Score 11.6; DB 6; Length 26;

Best Local Similarity 77.8%; Pred. No. 1.8e+04; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 3 gcgggtcacagtgtcg 20
||||| ||||| |||
Db 20 GCGGGCAACACGTCGTG 3

RESULT 10
US-09-915-060-9
; Sequence 9, Application US/09915060
; GENERAL INFORMATION:
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnol
; TITLE OF INVENTION: Novel internal ribosome entry site, vector containing same and th
; FILE REFERENCE: 2676-4976US
; CURRENT APPLICATION NUMBER: US/09/915,060
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 99200216.2
; PRIOR FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: in frame NotI
US-09-915-060-9

Query Match 58.0%; Score 11.6; DB 7; Length 30;
Best Local Similarity 77.8%; Pred. No. 1.9e+04;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgcgggtcacagtgtgc 19
||||| ||||| |||
Db 12 cgcggccgcagagtggac 29

RESULT 11
US-09-915-060-43
; Sequence 43, Application US/09915060
; GENERAL INFORMATION:
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnol
; TITLE OF INVENTION: Novel internal ribosome entry site, vector containing same and th
; FILE REFERENCE: 2676-4976US
; CURRENT APPLICATION NUMBER: US/09/915,060
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 99200216.2
; PRIOR FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-09-915-060-43

Query Match 58.0%; Score 11.6; DB 7; Length 30;
Best Local Similarity 77.8%; Pred. No. 1.9e+04;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgcgggtcacagtgtgc 19
||||| ||||| |||
Db 12 cgcggccgcagagtggac 29

RESULT 12
US-60-298-340-21/c
; Sequence 21, Application US/60298340
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN GROWTH FACTOR WITH HOMOL
; FILE REFERENCE: D0166 PSP
; CURRENT APPLICATION NUMBER: US/60/298,340
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 39
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-298-340-21

Query Match 58.0%; Score 11.6; DB 8; Length 39;
Best Local Similarity 77.8%; Pred. No. 1.9e+04;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcgggtcacagtgtcg 20
||||| ||||| |||
Db 25 GCTGTCTCACAGCGCGG 8

RESULT 13
US-09-787-252-17/c
; Sequence 17, Application US/09787252
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF RHO G EXPRESSION
; FILE REFERENCE: RTSP-0106
; CURRENT APPLICATION NUMBER: US/09/787,252
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 09/161,015
; PRIOR FILING DATE: 1998-09-25
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 17
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-09-787-252-17

Query Match 57.0%; Score 11.4; DB 7; Length 18;
Best Local Similarity 92.3%; Pred. No. 2.3e+04;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cgggtcacagtgc 16
||||| ||||| |||
Db 17 CGGGCGCACAGTG 5

RESULT 14
US-09-801-274-629
; Sequence 629, Application US/09801274
; GENERAL INFORMATION:
; APPLICANT: Cargill, Michele
; APPLICANT: Ireland, James S.
; TITLE OF INVENTION: HUMAN SINGLE NUCLEOTIDE POLYMORPHISMS
; FILE REFERENCE: 2825.2009-001
; CURRENT APPLICATION NUMBER: US/09/801,274
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 60/187,510
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 60/206,129
; PRIOR FILING DATE: 2000-05-22
; NUMBER OF SEQ ID NOS: 1802
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 629
```

; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-801-274-629

Query Match 57.0%; Score 11.4; DB 6; Length 31;  
Best Local Similarity 80.0%; Pred. No. 2.3e+04;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 6 gggtcacagtggctcg 20  
|| ||||| :|  
Db 3 gggtcacagtggcsg 17

RESULT 15  
US-09-735-271-1558/C  
; Sequence 1558, Application US/09735271  
; GENERAL INFORMATION:  
; APPLICANT: Daly, Mark J.  
; APPLICANT: Hudson, Thomas J.  
; APPLICANT: Lander, Eric S.  
; APPLICANT: Rioux, John  
; APPLICANT: Siminovitch, Kathy  
; TITLE OF INVENTION: IBD-RELATED POLYMORPHISMS  
; FILE REFERENCE: 2825.1025-002  
; CURRENT APPLICATION NUMBER: US/09/735,271  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 60/170,257  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 60/196,046  
; PRIOR FILING DATE: 2000-04-10  
; NUMBER OF SEQ ID NOS: 2058  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1558  
; LENGTH: 40  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(40)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-735-271-1558

Query Match 57.0%; Score 11.4; DB 7; Length 40;  
Best Local Similarity 92.3%; Pred. No. 2.4e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 gcggggtcacagt 15  
||||| :|||  
Db 35 GCGGGGTACAGT 23

Search completed: October 2, 2001, 16:54:58  
Job time: 17661 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:03:47 ; Search time 417.38 Seconds  
(without alignments)  
9.071 Million cell updates/sec

Title: US-09-757-100B-9

Perfect score: 20

Sequence: 1 ttttctagatgctaggtat 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 460742

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match		Query Match Length	DB ID	Description
		Match	Length			
1	20	100.0	20	3	US-09-377-310-9	Sequence 9, Appl
2	15	75.0	15	3	US-09-377-310-29	Sequence 29, Appl
3	15	75.0	20	3	US-09-377-310-8	Sequence 8, Appl
4	13.6	68.0	31	1	US-08-219-012-39	Sequence 39, Appl
5	13.6	68.0	31	4	US-08-687-421-227	Sequence 227, App
6	13	65.0	15	3	US-09-377-310-28	Sequence 28, Appl
7	12.2	61.0	29	3	US-08-792-832A-15	Sequence 15, Appl
8	12.2	61.0	29	3	US-09-033-055A-10	Sequence 10, Appl
9	12.2	61.0	35	4	US-08-686-968C-30	Sequence 28, Appl
10	12.2	61.0	35	4	US-08-686-968C-30	Sequence 30, Appl
11	12.2	61.0	35	4	US-08-686-968C-32	Sequence 32, Appl
12	12.2	61.0	35	4	US-08-686-968C-34	Sequence 34, Appl
13	12.2	61.0	41	2	US-08-478-386A-8	Sequence 8, Appl
14	12.2	61.0	41	2	US-08-292-597-8	Sequence 8, Appl
15	12.2	61.0	41	2	US-08-388-653-8	Sequence 8, Appl
16	12.2	61.0	41	2	US-08-473-985-8	Sequence 8, Appl
17	12.2	61.0	41	2	US-08-483-898-8	Sequence 8, Appl
18	12.2	61.0	41	3	US-09-087-716-8	Sequence 8, Appl
19	12.2	61.0	41	3	US-09-157-753-8	Sequence 8, Appl
20	12.2	61.0	41	3	US-09-157-230-8	Sequence 8, Appl
21	12.2	61.0	41	3	US-09-087-811-8	Sequence 8, Appl
22	12.2	61.0	41	3	US-09-156-855-8	Sequence 8, Appl
23	12.2	61.0	41	3	US-09-158-010-8	Sequence 8, Appl
24	12.2	61.0	41	4	US-09-087-647-8	Sequence 8, Appl
25	12	60.0	35	1	US-08-810-116-2	Sequence 2, Appl
26	12	60.0	35	2	US-07-930-548A-2	Sequence 2, Appl
27	12	60.0	35	2	US-08-818-604-22	Sequence 22, Appl

c

ALIGNMENTS

RESULT 1

US-09-377-310-9  
; Sequence 9, Application US/09377310B  
; Patent No. 6133031

GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.

; APPLICANT: Gaarde, William A.

; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase

; TITLE OF INVENTION: Expression

; FILE REFERENCE: ISPH-0389

; CURRENT APPLICATION NUMBER: US/09/377,310B

; CURRENT FILING DATE: 1999-08-19

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 20

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: antisense sequence

US-09-377-310-9

Query Match 100.0%; Score 20; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.041;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttctagatgctaggtat 20

Db 1 ttttctagatgctaggtat 20

RESULT 2

US-09-377-310-29  
; Sequence 29, Application US/09377310B  
; Patent No. 6133031

GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.

; APPLICANT: Gaarde, William A.

; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase

; TITLE OF INVENTION: Expression

; FILE REFERENCE: ISPH-0389

; CURRENT APPLICATION NUMBER: US/09/377,310B

; CURRENT FILING DATE: 1999-08-19

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 29

; LENGTH: 15

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-29

Query Match 75.0%; Score 15; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ttgctagatgctagg 17  
Db 1 ttgctagatgctagg 15

RESULT 3  
US-09-377-310-8  
; Sequence 8, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-8

Query Match 75.0%; Score 15; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ctagatgctaggat 20  
Db 1 ctagatgctaggat 15

RESULT 4  
US-08-219-012-39  
; Sequence 39, Application US/08219012  
; Patent No. 5543293  
; GENERAL INFORMATION:  
; APPLICANT: Larry Gold  
; APPLICANT: Diane Tasset  
; TITLE OF INVENTION: Ligands of Thrombin  
; NUMBER OF SEQUENCES: 92  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Beaton & Swanson, P.C.  
; STREET: 4582 South Ulster Street Parkway, Suite #  
; STREET: 403  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80237  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/219,012  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION NUMBER: none

; ATTORNEY/AGENT INFORMATION:  
; NAME: Barry J. Swanson  
; REGISTRATION NUMBER: 33,215  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 850-9900  
; TELEFAX: (303) 850-9401  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 31 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-219-012-39

Query Match 68.0%; Score 13.6; DB 1; Length 31;  
Best Local Similarity 80.0%; Pred. No. 98;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tttagctagatgctaggat 20  
Db 2 TTTTGGTATAGCTAGGTGT 21

RESULT 5  
US-08-687-421-227  
; Sequence 227, Application US/08687421  
; Patent No. 6177557  
; GENERAL INFORMATION:  
; APPLICANT: Gold, Larry  
; APPLICANT: Janjic, Nebojsa  
; APPLICANT: Tasset, Diane  
; TITLE OF INVENTION: HIGH-AFFINITY LIGANDS OF BASIC  
; TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR AND  
; TITLE OF INVENTION: THROMBIN  
; NUMBER OF SEQUENCES: 445  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Swanson & Bratschun, L.L.C.  
; STREET: 8400 E. Prentice Avenue, Suite 200  
; CITY: Englewood  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/687,421  
; FILING DATE: 08-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/195,005  
; FILING DATE: 10-FEBRUARY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE: 22-APRIL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/219,012  
; FILING DATE: 28-MARCH-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/973,333  
; FILING DATE: 11-NOVEMBER-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/714,131  
; FILING DATE: 10-JUNE-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/536,428  
; FILING DATE: 11-JUNE-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barry J. Swanson



; REGISTRATION NUMBER: 33,215  
; REFERENCE/DOCKET NUMBER: NEX07/PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 793-3333  
; TELEFAX: (303) 793-3433  
; INFORMATION FOR SEQ ID NO: 227:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 31 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-687-421-227

Query Match 68.0%; Score 13.6; DB 4; Length 31;  
Best Local Similarity 80.0%; Pred. No. 98;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttttgcgtagctaggtat 20  
||||| ||| ||||| |||  
Db 2 TTTTGGTATAGGCTAGGTGT 21

## RESULT 6

US-09-377-310-28  
; Sequence 28, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:

; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 15

; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
; US-09-377-310-28

Query Match 65.0%; Score 13; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 agatgctaggtat 20  
||||| ||||| ||||| |||  
Db 1 agatgctaggtat 13

## RESULT 7

US-08-792-832A-15  
; Sequence 15, Application US/08792832A  
; Patent No. 6017734  
; GENERAL INFORMATION:

; APPLICANT: Summers Dr., Max D.  
; APPLICANT: Braunagel Dr., Sharon C.  
; APPLICANT: Hong Dr., Tao  
; TITLE OF INVENTION: UNIQUE NUCLEOTIDE AND AMINO ACID  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,832A  
; FILING DATE: 30-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/000,955  
; FILING DATE: 07-JUL-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/678,435  
; FILING DATE: 03-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hibler, David W.  
; REGISTRATION NUMBER: 41,071  
; REFERENCE/DOCKET NUMBER: TAMK:190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 512/418-3000  
; TELEFAX: 512/474-7577  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 29 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-792-832A-15

Query Match 61.0%; Score 12.2; DB 3; Length 29;  
Best Local Similarity 82.4%; Pred. No. 5.3e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttgctagatgctaggtat 19  
||||| ||| |||||  
Db 13 TTGCTATTGTATAGGTA 29

## RESULT 8

US-09-033-055A-10  
; Sequence 10, Application US/09033055A  
; Patent No. 6069241  
; GENERAL INFORMATION:

; APPLICANT: OHKAWA, HIDEO  
; APPLICANT: IMAISHI, HIROMASA  
; TITLE OF INVENTION: NOVEL CYTOCHROME P450 GENE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PILLSBURY, MADISON & SUTRO  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/033,055A  
; FILING DATE: MARCH 2, 1998  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PAUL E. WHITE, JR.  
; REGISTRATION NUMBER: 32,011  
; REFERENCE/DOCKET NUMBER: 9437/251563  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3651  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 29

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-09-033-055A-10

Query Match 61.0%; Score 12.2; DB 3; Length 29;

Best Local Similarity 82.4%; Pred. No. 5.3e+02;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttgctagatgctaggta 19

||||| |||||||

Db 2 TTGCTATTGCTAGGTA 18

RESULT 9

US-08-686-968C-28

; Sequence 28, Application US/08686968C

; Patent No. 6221361

; GENERAL INFORMATION:

; APPLICANT: Cochran, Mark D.

; APPLICANT: Junker, David E.

; TITLE OF INVENTION: Recombinant Swinepox Virus

; FILE REFERENCE: 39119-H/JML

; CURRENT APPLICATION NUMBER: US/08/686,968C

; CURRENT FILING DATE: 1996-07-25

; NUMBER OF SEQ ID NOS: 231

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 28

; LENGTH: 35

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: primer

US-08-686-968C-28

Query Match

Best Local Similarity 61.0%; Score 12.2; DB 4; Length 35;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctagatgctagggt 18

||||| |||||||

Db 18 ttctctaggtgctagtt 34

RESULT 10

US-08-686-968C-30

; Sequence 30, Application US/08686968C

; Patent No. 6221361

; GENERAL INFORMATION:

; APPLICANT: Cochran, Mark D.

; APPLICANT: Junker, David E.

; TITLE OF INVENTION: Recombinant Swinepox Virus

; FILE REFERENCE: 39119-H/JML

; CURRENT APPLICATION NUMBER: US/08/686,968C

; CURRENT FILING DATE: 1996-07-25

; NUMBER OF SEQ ID NOS: 231

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 30

; LENGTH: 35

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: primer

US-08-686-968C-30

Query Match

Best Local Similarity 61.0%; Score 12.2; DB 4; Length 35;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctagatgctagggt 18

||||| |||||||

Db 18 ttctctaggtgctagtt 34

RESULT 11

US-08-686-968C-32

; Sequence 32, Application US/08686968C

; Patent No. 6221361

; GENERAL INFORMATION:

; APPLICANT: Cochran, Mark D.

; APPLICANT: Junker, David E.

; TITLE OF INVENTION: Recombinant Swinepox Virus

; FILE REFERENCE: 39119-H/JML

; CURRENT APPLICATION NUMBER: US/08/686,968C

; CURRENT FILING DATE: 1996-07-25

; NUMBER OF SEQ ID NOS: 231

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 32

; LENGTH: 35

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: primer

US-08-686-968C-32

Query Match

Best Local Similarity 61.0%; Score 12.2; DB 4; Length 35;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctagatgctagggt 18

||||| |||||||

Db 18 ttctctaggtgctagtt 34

RESULT 12

US-08-686-968C-34

; Sequence 34, Application US/08686968C

; Patent No. 6221361

; GENERAL INFORMATION:

; APPLICANT: Cochran, Mark D.

; APPLICANT: Junker, David E.

; TITLE OF INVENTION: Recombinant Swinepox Virus

; FILE REFERENCE: 39119-H/JML

; CURRENT APPLICATION NUMBER: US/08/686,968C

; CURRENT FILING DATE: 1996-07-25

; NUMBER OF SEQ ID NOS: 231

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 34

; LENGTH: 35

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: primer

US-08-686-968C-34

Query Match

Best Local Similarity 61.0%; Score 12.2; DB 4; Length 35;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctagatgctagggt 18

||||| |||||||

Db 18 ttctctaggtgctagtt 34

RESULT 13

US-08-478-386A-8

; Sequence 8, Application US/08478386A

; Patent No. 5830462

GENERAL INFORMATION:  
APPLICANT: Crabtree, Gerald R.  
APPLICANT: Schreiber, Stuart L.  
APPLICANT: Wandless, Thomas J.  
APPLICANT: Spencer, David M.  
APPLICANT: Wandless, Peter  
APPLICANT: Belshaw, Thomas J.  
TITLE OF INVENTION: REGULATED TRANSCRIPTION OF TARGETED  
GENES AND OTHER BIOLOGICAL EVENTS  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
STREET: 26 Landsdowne Street  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02139  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC/DOS/MS/DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,386A  
FILING DATE: 07/JUN/1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Figg, E. Anthony  
REGISTRATION NUMBER: 27,195  
REFERENCE/DOCKET NUMBER: 2054-114A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 783-6040  
TELEFAX: (202) 783-6031  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 41 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6..11  
OTHER INFORMATION: /note= "Xho I restriction site."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 12..41  
OTHER INFORMATION: /note= "Region of homology with  
target sequence."  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9..41  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 28  
OTHER INFORMATION: /note= "A to G."  
US-08-478-386A-8

Query Match 61.0%; Score 12.2; DB 2; Length 41;  
Best Local Similarity 82.4%; Pred. No. 5.5e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctagatgctaggt 18  
| ||||| |||||  
Db 13 TCTGCTACTTGCTAGGT 29

RESULT 14  
US-08-292-597-8  
Sequence 8, Application US/08292597  
Patent No. 5834266  
GENERAL INFORMATION:  
APPLICANT: Gerald R. Crabtree

APPLICANT: Schreiber, Stuart L.  
APPLICANT: Spencer, David M.  
APPLICANT: Wandless, Thomas J.  
APPLICANT: Belshaw, Peter  
TITLE OF INVENTION: Regulated Apoptosis  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
STREET: 26 Landsdowne Street  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02139  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC/DOS/MS/DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,597  
FILING DATE: 18/AUG/1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Figg, E. Anthony  
REGISTRATION NUMBER: 27,195  
REFERENCE/DOCKET NUMBER: 2054-108A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 783-6040  
TELEFAX: (202) 783-6031  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 41 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6..11  
OTHER INFORMATION: /note= "Xho I restriction site."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 12..41  
OTHER INFORMATION: /note= "Region of homology with  
target sequence."  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9..41  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 28  
OTHER INFORMATION: /note= "A to G."  
US-08-292-597-8

Query Match 61.0%; Score 12.2; DB 2; Length 41;  
Best Local Similarity 82.4%; Pred. No. 5.5e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctagatgctaggt 18  
| ||||| |||||  
Db 13 TCTGCTACTTGCTAGGT 29

RESULT 15  
US-08-388-653-8  
Sequence 8, Application US/08388653  
Patent No. 5869337  
GENERAL INFORMATION:  
APPLICANT: Crabtree, Gerald R.

APPLICANT: Schreiber, Stuart L.  
APPLICANT: Spencer, David M.  
APPLICANT: Wandless, Thomas J.  
APPLICANT: Belshaw, Peter  
TITLE OF INVENTION: REGULATED TRANSCRIPTION OF TARGETED  
TITLE OF INVENTION: GENES AND OTHER BIOLOGICAL EVENTS  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
STREET: 26 Landsdowne Street  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02139  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC/DOS/MS/DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/388,653  
FILING DATE: 14-FEB-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/478,386  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Figg, E. Anthony  
REGISTRATION NUMBER: 27,195  
REFERENCE/DOCKET NUMBER: 2054-114A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 783-6040  
TELEFAX: (202) 783-6031  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 41 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6..11  
OTHER INFORMATION: /note= "Xho I restriction site."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 12..41  
OTHER INFORMATION: /note= "Region of homology with  
OTHER INFORMATION: target sequence."  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9..41  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 28  
OTHER INFORMATION: /note= "A to G."  
US-08-388-653-8

Query Match 61.0%; Score 12.2; DB 2; Length 41;  
Best Local Similarity 82.4%; Pred. No. 5.5e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ttgtctagatgctaggt 18  
| | | | | | | | | |  
Db 13 TCTGCTACTTGCTAGGT 29

Search completed: October 2, 2001, 16:03:48  
Job time: 14592 sec

new anticancer compounds for inhibiting local destruction kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -  
XX

PS Claim 3; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
 CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
 CC is involved in integrin-mediated signal transduction, and is implicated  
 CC in cancer, particularly colon, breast and oral tumours, embryonic  
 CC development disorders, angiogenic disorders and wound healing. The  
 CC antisense sequences, including the one shown here, can be used in the  
 CC treatment of all of these.

XX Sequence 20 BP; 4 A; 2 C; 5 G; 9 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 0.16;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgtagtagtagtat 20  
 |||||

Db 1 ttgtgtagtagtagtat 20

#### RESULT 2

AAC65561  
 ID AAC65561 standard; DNA; 15 BP.

AC AAC65561;

DT 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #27.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
 KW embryonic development disorder; angiogenic disorder; wound healing;  
 KW antisense; phosphorothioate; ss.

OS Homo sapiens.

XX US6133031-A.

PD 17-OCT-2000.

XX 19-AUG-1999; 99US-0377310.

PR 19-AUG-1999; 99US-0377310.

XX (ISIS-) ISIS PHARM INC.

PI Monia BP, Gaarde WA;

XX WPI; 2001-006141/01.

PT New antisense compounds for inhibiting focal adhesion kinase  
 PT expression, especially useful for inhibiting retinal  
 PT neovascularization, or for diagnosing and treating e.g. colon cancer -

XX Example 2; Column 25; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
 CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
 CC is involved in integrin-mediated signal transduction, and is implicated  
 CC in cancer, particularly colon, breast and oral tumours, embryonic  
 CC development disorders, angiogenic disorders and wound healing. The  
 CC antisense sequences, including the one shown here, can be used in the  
 CC treatment of all of these.

XX Sequence 15 BP; 3 A; 2 C; 5 G; 5 T; 0 other;

Query Match 75.0%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ttgtagtagtagg 17  
 |||||

Db 1 ttgtagtagtagg 15

#### RESULT 3

AAC65540  
 ID AAC65540 standard; DNA; 20 BP.

XX AAC65540;

DT 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #6.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
 KW embryonic development disorder; angiogenic disorder; wound healing;  
 KW antisense; phosphorothioate; ss.

OS Homo sapiens.

XX US6133031-A.

PD 17-OCT-2000.

XX 19-AUG-1999; 99US-0377310.

PR 19-AUG-1999; 99US-0377310.

XX (ISIS-) ISIS PHARM INC.

PI Monia BP, Gaarde WA;

XX WPI; 2001-006141/01.

PT New antisense compounds for inhibiting focal adhesion kinase  
 PT expression, especially useful for inhibiting retinal  
 PT neovascularization, or for diagnosing and treating e.g. colon cancer -

XX Claim 3; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
 CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
 CC is involved in integrin-mediated signal transduction, and is implicated  
 CC in cancer, particularly colon, breast and oral tumours, embryonic  
 CC development disorders, angiogenic disorders and wound healing. The  
 CC antisense sequences, including the one shown here, can be used in the  
 CC treatment of all of these.

XX Sequence 20 BP; 4 A; 4 C; 5 G; 7 T; 0 other;

Query Match 75.0%; Score 15; DB 22; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ctatagtagtagtat 20  
 |||||

Db 1 ctatagtagtagtat 15

#### RESULT 4

AAX33269  
 ID AAX33269 standard; DNA; 20 BP.

XX AAX33269;

DT 30-JUN-1999 (first entry)

XX PEP2 alpha A gene expression regulating DNA PCR primer SEQ ID NO:26.

XX PEP2 alpha A gene; expression; regulation; bone disease;

KW osteoporosis; PCR primer; ss.

```
xx OS Synthetic.
xx PN WO9911787-A1.
xx PD 11-MAR-1999.
xx PF 02-SEP-1998; 98WO-JP03920.
xx PR 08-APR-1998; 98JP-0114135.
xx PR 02-SEP-1997; 97JP-0254250.
xx PR 15-OCT-1997; 97JP-0299407.
xx PA (SUMU ) SUMITOMO PHARM CO LTD.
xx PI Fujiwara M, Harada H, Katsumata T, Nakatsuka M;
xx PI Ogawa S, Tagashira S;
xx DR WPI; 1999-243621/20.
xx CC DNA regulating expression of PEBP2 alphaA gene to produce regulator
xx PT protein, useful as promoter for prevention or/and treatment of bone
xx PT diseases e.g. osteoporosis
xx PS Example 6; Page 39; 118pp; Japanese.
xx CC The present invention describes DNA which participates in the regulation
xx CC of expression of PEBP2 alpha A gene. The DNA produces a regulator
xx CC protein with the activity of promoting bone formation and can serve as a
xx CC promoter for prevention and treatment of bone diseases including
xx CC osteoporosis. The present sequence represents a PCR primer used in an
xx CC example from the present invention.
xx CC Sequence 20 BP; 6 A; 1 C; 6 G; 7 T; 0 other;

Query Match 69.0%; Score 13.8; DB 20; Length 20;
Best Local Similarity 88.2%; Pred. No. 2.6e+02;
. Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgtctagatgctaggt 18
Db ||||| ||||| |||
3 ttgtataaatgctaggt 19

RESULT 5
AA48327/C
ID AAA48327 standard; DNA; 36 BP.
AC AAA48327;
XX 28-SEP-2000 (first entry)
DE HIV gp140 gene PCR primer, gp140CysEcoRI.
XX Antigen presentation; vaccine; infectious disease; allergy; cancer;
KW molecular scaffold; immune response; farm animal; organiser; gp140;
KW immunostimulatory; cytostatic; antiallergy; human immunodeficiency virus;
KW HIV; PCR primer; ss.
XX Human immunodeficiency virus.
OS OS
XX WO200032227-A2.
XX 08-JUN-2000.
PD 30-NOV-1999; 99WO-IB01925.
XX 30-NOV-1998; 98US-0110414.
PR 08-JUL-1999; 99US-0142788.
XX (CYTO-) CYTOS BIOTECHNOLOGY AG.
PA XX
```

```
PI Renner WA, Hennecke F, Nieba L, Bachmann M;
XX WPI; 2000-412159/35.
XX Composition for use as vaccine against infectious diseases and in
PT treatment of cancer and allergies comprises non-naturally occurring
PT molecular scaffold and antigen or antigenic determinant -
XX Example 26; Page 77; 102pp; English.
XX A new method for developing vaccines has been identified, in which a
CC non-naturally occurring molecular scaffold, having a core particle and a
CC covalently attached organiser, is attached to an antigen or antigenic
CC determinant. The scaffold and antigen or antigenic determinant interact
CC to form an ordered and repetitive antigen array. The composition is
CC useful as a vaccine against infectious diseases, to induce immune
CC responses in farm animals and also in the treatment of cancer and
CC allergies. The present sequence is the PCR primer, gp140CysEcoRI. This
CC primer was used to amplify the coding sequence for the gp140 gene of
CC human immunodeficiency virus (HIV). The PCR product was then used as an
CC antigen in the present invention.
XX CC Sequence 36 BP; 11 A; 10 C; 6 G; 9 T; 0 other;

Query Match 69.0%; Score 13.8; DB 21; Length 36;
Best Local Similarity 88.2%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tgctagatgctaggtat 20
Db ||||| ||||| ||
22 TGCTAGCTGCTAGGAAT 6

RESULT 6
AAT00213
ID AAT00213 standard; DNA; 31 BP.
XX AAT00213;
AC AAT00213;
XX 14-AUG-1996 (first entry)
DT Thrombin DNA ligand, clone #17.
XX Family 1; family 2; ligand: thrombin;
KW systematic evolution of ligands by exponential enrichment; SELEX;
KW heparin; selection; region of homology; inhibitor; ss.
XX Synthetic.
OS OS
XX WO9521853-A1.
XX 17-AUG-1995.
PD 06-FEB-1995; 95WO-US01458.
XX 28-MAR-1994; 94US-0219012.
PR 10-FEB-1994; 94US-0195005.
PR 11-JUN-1990; 90US-0536428.
PR 10-JUN-1991; 91US-0714131.
PR 22-APR-1993; 93US-0061691.
XX (NEXS-) NEXSTAR PHARM INC.
XX Gold L, Janjic N, Tasset D;
PI WPI; 1995-293073/38.
XX Identification of ligands to basic fibroblast growth factor and
PT thrombin - which can be modified for increased in vivo stability
XX Claim 39; Page 95; 236pp; English.
XX
```

CC The sequences given in AAT00202-25 and AAT00227-57 represent two groups  
 CC of ligands to thrombin. These sequences were isolated using the single  
 CC stranded DNA molecules given in AAT00201 and AAT00226 which comprise a  
 CC 30N and a 60N variable region, respectively. These ligands were  
 CC isolated using systematic evolution of ligands by exponential enrichment  
 CC (SELEX). The selection was conducted in a buffer solution at 37 deg. C.  
 CC After 12 rounds of selection, no additional improvement in binding was  
 CC seen. By studying regions of homology between the isolated ligands, a  
 CC truncated ligand of 38 nucleotides (see AAO98403-04) was identified which  
 CC retains high affinity binding and inhibits clotting. These ligands are  
 CC inhibitors of thrombin and are therefore useful in treating thrombin  
 CC mediated conditions and in studying the structure and binding of  
 CC thrombin.

XX Sequence 31 BP; 5 A; 3 C; 11 G; 12 T; 0 other;

Query Match 68.0%; Score 13.6; DB 16; Length 31;  
 Best Local Similarity 80.0%; Pred. No. 3.5e+02;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ttttctagatgctaggtat 20  
 ||||| || | ||||| |  
 Db 2 ttttggtatagctaggtgt 21

RESULT 7  
 AAF70765  
 ID AAF70765 standard; DNA; 31 BP.

XX AC AAF70765;

XX DT 20-APR-2001 (first entry)

XX DE Thrombin high affinity ligand #12.

XX KW Ligand; basic fibroblast growth factor; bFGF; gene therapy; vascular;  
 KW atherosclerosis; angioplasty; stability; ss.

XX OS Unidentified.

XX PN US6177557-B1.

XX PD 23-JAN-2001.

XX PF 05-AUG-1996; 96US-0687421.

XX PR 11-JUN-1990; 90US-0536428.

XX PR 10-JUN-1991; 91US-0714131.

XX PR 06-NOV-1992; 92US-0973333.

XX PR 10-FEB-1994; 94US-0195005.

XX PR 28-MAR-1994; 94US-0219012.

XX PA (NEXS-) NEXSTAR PHARM INC.

XX PI Janjic N, Gold L, Tasset D;

XX DR WPI; 2001-158583/16.

XX PS Example 19; Column 57-58; 153pp; English.

XX CC The present invention relates to a purified and isolated non-naturally  
 CC occurring DNA ligands to basic fibroblast growth factor (bFGF).

XX CC The ligands are useful as part of gene therapy treatments and  
 CC for diagnosing pathogenesis of vascular diseases including

XX CC initiation and progression of atherosclerosis, acute coronary  
 CC syndromes, vein graft disease and restenosis following coronary

XX CC angioplasty. The ligands have improved stability in vivo.

SQ Sequence 31 BP; 5 A; 3 C; 11 G; 12 T; 0 other;

Query Match 68.0%; Score 13.6; DB 22; Length 31;  
 Best Local Similarity 80.0%; Pred. No. 3.5e+02;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ttttctagatgctaggtat 20  
 ||||| || | ||||| |  
 Db 2 ttttggtatagctaggtgt 21

RESULT 8

AAZ93950

ID AAZ93950 standard; DNA; 27 BP.

XX AC AAZ93950;

XX DT 29-AUG-2000 (first entry)

XX DE CaMV Calibrator probe sequence used in detection method.

XX KW Detection; identification; PCR; polymerase chain reaction; probe;  
 KW primer; calibrator; diagnosis; pathogen; virus; bacteria; HHV-6;  
 KW HHV-7; HHV-8; human herpes virus; human immunodeficiency virus; HIV;  
 KW AIDS; acquired immune deficiency syndrome;  
 KW Mycobacterium tuberculosis; ss.

XX OS Synthetic.

XX PN WO200029613-A1.

XX PD 25-MAY-2000.

XX PF 17-NOV-1999; 99WO-EP08847.

XX PR 17-NOV-1998; 98IT-MI02491.

XX PA (SANR-) FOND CENT SAN RAFFAELE DEL MONTE TABOR.

XX PI Locatelli G, Lusso P, Malnati M, Salvatori F, Scarlatti G;

XX DR WPI; 2000-387819/33.

XX PT Quantitative detection of nucleic acids, based on the use of a  
 PT calibrator, suitable primers and probes, and a nucleic acid polymerase  
 PT with 5'-3' nuclease activity

XX Example 1; Page 13; 39pp; English.

XX CC A new method for the quantitative detection of nucleic acids in a  
 CC sample which uses a calibrator, suitable primers and probes, and  
 CC a nucleic acid polymerase with 5'-3' nuclease activity is described.  
 CC The method comprises adding a calibrator nucleotide to the sample,  
 CC the calibrator nucleotide having the same sequence as the target with  
 CC the exception of one or more regions which in the target hybridize  
 CC with a probe labeled with a reporter and quencher, or which hybridize  
 CC with the probe and with 2 or more primers. These regions have a  
 CC different, randomized nucleotide sequence and a similar Tm. Then  
 CC extracting the calibrator and target nucleotides from the sample.  
 CC Probes and primers are then added to the extracted sample/calibrator  
 CC mixture and PCR is performed. Reaction in the presence of the target  
 CC nucleic acid specific probe permits quantitation of the copy number  
 CC of the extracted target nucleic acid. The reaction in the presence of  
 CC the calibrator nucleotide permits quantitation of the calibrator copy  
 CC number. The reaction in the presence of both permits calculation of  
 CC the total number of target templates and calibrator, allowing  
 CC calculation of the percentage of calibrator recovery. The method may  
 CC be used for the diagnosis of viral and/or other pathogenic agents in  
 CC body fluids, and to monitor safety and/or genetic composition of  
 CC waters, foods, and plant species used in the alimentary field. The  
 CC method is used to quantitatively detect the genomic nucleic acid of  
 CC human herpes virus (HHV)-6, HHV-7, HHV-8 and human immunodeficiency



CC virus (HIV). The method provides an enhanced sensitivity, accuracy,  
CC and precision and a reduced measure viability. For specific probe,  
CC primer and calibrator sequences used in the method, see GENESEQ  
CC records AA293935-293965.

XX SQ Sequence 27 BP; 5 A; 7 C; 7 G; 8 T; 0 other;  
Query Match 66.0%; Score 13.2; DB 21; Length 27;  
Best Local Similarity 83.3%; Pred. No. 5.6e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttgctagatgctaggtat 20  
| | | | | | | | | | | |  
Db 2 tcgctacatgctagcat 19

RESULT 9  
AA79014/c  
ID AAA79014 standard; DNA; 31 BP.  
XX AC AAA79014;  
XX DT 20-NOV-2000 (first entry)  
XX DE Human genomic DNA polymorphic site sequence tag SEQ ID NO:384.  
XX KW Human; genomic DNA; polymorphism; genome; allele-specific; primer;  
KW probe; hybridisation; polymorphic site; forensic; paternity testing;  
KW medicine; phenotypic trait; genetic analysis; genetic mapping; ds.  
XX OS Homo sapiens.

XX EP1024200-A2.  
XX PF 02-AUG-2000.  
XX PR 26-JAN-2000; 2000EP-0250023.  
XX PI 27-JAN-1999; 99US-0238402.  
XX PA (AFFY-) AFFYMETRIX INC.  
XX PI Patil N, Shah N, Warrington JA;  
XX WPI; 2000-500198/45.

XX Human genomic polymorphic nucleic acid segments, allele specific  
PT primers and probes, and methods of analysis, useful for e.g. forensics,  
PT paternity testing, genetic mapping,  
XX Claim 1; Page 15; 14pp; English.

XX The present invention describes a nucleic acid segment of 10-100  
CC contiguous bases chosen from one of 632 fragments (AAA78631 to  
CC AAA79262), where the segment comprises a polymorphic site or an  
CC immediately adjacent base, or the complement of the segment. Also  
CC described are: (1) an allele-specific oligonucleotide that hybridises to  
CC a segment of the novelty; (2) an isolated nucleic acid comprising a  
CC sequence of the novelty where the polymorphic site within the sequence is  
CC occupied by a base other than the reference base indicated in the  
CC specification; and (3) analysing a nucleic acid, comprising obtaining a  
CC nucleic acid from an individual, and determining a base occupying any one  
CC of the polymorphic sites of the novelty. The nucleic acid segments and  
CC method can be used to analyse an individual's nucleic acid sequences for  
CC the presence of polymorphisms. The method can also be used to test for a  
CC disease phenotype and correlate the presence of the phenotype with a  
CC particular polymorphism. The presence of polymorphic sites are useful  
CC for, e.g. forensics, paternity testing, correlation of polymorphisms  
CC with phenotypic traits and for genetic mapping of phenotypic traits.  
CC AAA78631 to AAA79262 represent sequence tags of human genomic DNA  
CC fragments containing polymorphic sites. The base occupying the  
CC polymorphic site is indicated using IUPAC-IUB nomenclature.

XX SQ Sequence 31 BP; 9 A; 8 C; 8 G; 5 T; 1 other;  
Query Match 66.0%; Score 13.2; DB 21; Length 31;  
Best Local Similarity 75.0%; Pred. No. 5.7e+02;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttgtgctagatgctaggtat 20  
| | | | | | | | | | | |  
Db 24 TCTGCTCRATGCTAGAGAT 5

RESULT 10  
AAC65560  
ID AAC65560 standard; DNA; 15 BP.  
XX AC AAC65560;  
XX DT 12-FEB-2001 (first entry)  
XX DE Human focal adhesion kinase antisense sequence #26.  
XX KW Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.  
XX OS Homo sapiens.

XX PN US6133031-A.  
XX PD 17-OCT-2000.  
XX PF 19-AUG-1999; 99US-0377310.  
XX PR 19-AUG-1999; 99US-0377310.  
XX PA (ISIS-) ISIS PHARM INC.  
XX PI Monia BP, Gaarde WA;  
XX DR WPI; 2001-006141/01.  
XX PT New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -  
XX Example 2; Column 25; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, embryonic  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.  
XX SQ Sequence 15 BP; 4 A; 2 C; 4 G; 5 T; 0 other;

Query Match 65.0%; Score 13; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 agatgctaggtat 20  
| | | | | | | | | | | |  
Db 1 agatgctaggtat 13  
RESULT 11  
AA299288  
ID AA299288 standard; DNA; 36 BP.  
XX

```

AC AA299288;
XX
XX 03-JUL-2000 (first entry)
XX
DE Nucleotide sequence of a clone from a RNA-protein fusion library.
XX
XX RNA-protein fusion; protein-protein interaction; drug target;
KW protein chip; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 18..19
FT /*tag= a
FT /note= "an unspecified number of bases are present
FT between these nucleotides"
XX
XX WO200009737-A1.
XX
XX 24-FEB-2000.
XX
XX 16-AUG-1999; 99WO-US18603.
XX
XX 17-AUG-1998; 98US-0096818.
XX
XX (PHYL-) PHYLUS INC.
XX
XX Hammond PW, Lipovsek D;
XX
XX WPI; 2000-224362/19.
XX
XX Producing nucleic acids lacking 3'-untranslated regions to optimise
XX production of fusion proteins, used to produce fusion libraries for
XX identification of protein:protein interactions and drug targets for the
XX production of protein chips -
XX
XX Example 1; Fig 14; 52pp; English.
XX
XX The present sequence represents a fragment from a clone of a
XX RNA-protein fusion library comprising KIAA0111. specification describes
XX a method for removing the 3'-untranslated region of a DNA molecule
XX comprising an open reading frame (ORF). The method comprises providing
XX a DNA molecule having an ORF and a 3'-untranslated region, the DNA
XX having a 5' overhang and a blunt end at the 3' end, and treating the
XX DNA molecule with a 3'-5' exonuclease followed by a single-stranded
XX nuclease to remove the 3'-untranslated region. The products can be
XX used to produce RNA-protein fusion libraries. The fusion libraries
XX can be used for the identification of protein-protein interactions,
XX identification of drug targets, and hybridisation to solid supports to
XX create protein chips (or beads). The RNA-protein molecules may be
XX arranged in spatially defined arrays on the protein chips to carry out
XX large scale screening for protein or compound identification.
XX
XX Sequence 36 BP; 6 A; 9 C; 10 G; 11 T; 0 other;
XX
XX Query Match 63.0%; Score 12.6; DB 21; Length 36;
XX Best Local Similarity 78.9%; Pred. No. 1.2e+03;
XX Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgctagctagtgta 19
Db 5 ttttgatgaagctagta 23

RESULT 12
AAZ99289
ID AAZ99289 standard; DNA; 36 BP.
XX
XX AAZ99289;
XX
XX 03-JUL-2000 (first entry)
XX

```

```

DE Nucleotide sequence of a clone from a RNA-protein fusion library.
XX
XX RNA-protein fusion; protein-protein interaction; drug target;
KW protein chip; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 18..19
FT /*tag= a
FT /note= "an unspecified number of bases are present
FT between these nucleotides"
XX
XX WO200009737-A1.
XX
XX 24-FEB-2000.
XX
XX 16-AUG-1999; 99WO-US18603.
XX
XX 17-AUG-1998; 98US-0096818.
XX
XX (PHYL-) PHYLUS INC.
XX
XX Hammond PW, Lipovsek D;
XX
XX WPI; 2000-224362/19.
XX
XX Producing nucleic acids lacking 3'-untranslated regions to optimise
XX production of fusion proteins, used to produce fusion libraries for
XX identification of protein:protein interactions and drug targets for the
XX production of protein chips -
XX
XX Example 1; Fig 14; 52pp; English.
XX
XX The present sequence represents a fragment from a clone of a
XX RNA-protein fusion library comprising KIAA0111. specification describes
XX a method for removing the 3'-untranslated region of a DNA molecule
XX comprising an open reading frame (ORF). The method comprises providing
XX a DNA molecule having an ORF and a 3'-untranslated region, the DNA
XX having a 5' overhang and a blunt end at the 3' end, and treating the
XX DNA molecule with a 3'-5' exonuclease followed by a single-stranded
XX nuclease to remove the 3'-untranslated region. The products can be
XX used to produce RNA-protein fusion libraries. The fusion libraries
XX can be used for the identification of protein-protein interactions,
XX identification of drug targets, and hybridisation to solid supports to
XX create protein chips (or beads). The RNA-protein molecules may be
XX arranged in spatially defined arrays on the protein chips to carry out
XX large scale screening for protein or compound identification.
XX
XX Sequence 36 BP; 6 A; 9 C; 10 G; 11 T; 0 other;
XX
XX Query Match 63.0%; Score 12.6; DB 21; Length 36;
XX Best Local Similarity 78.9%; Pred. No. 1.2e+03;
XX Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgctagctagtgta 19
Db 5 ttttgatgaagctagta 23

RESULT 13
AAV99615/C
ID AAV99615 standard; DNA; 27 BP.
XX
XX AAV99615;
XX
XX 29-MAR-1999 (first entry)
XX
XX Maize rpoB gene primer rpoB#3.
XX
XX Promoter; nuclear encoded plastid RNA polymerase; NEP;
KW rpoB; chloroplast; transgenic plant; maize; PCR; primer; ss.

```

```

XX Synthetic.
OS Zea mays.
XX WO9855595-A1.
XX WO9855595-A1.
XX 10-DEC-1998.
XX 03-JUN-1998; 98WO-US11437.
XX 12-SEP-1997; 97US-0058670.
XX 03-JUN-1997; 97US-0048376.
XX (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX Maliga P, Silhavy D, Sritaman P;
XX WPI; 1999-070262/06.
XX Isolated nuclear-encoded plastid RNA polymerase promoter sequences -
PT useful for expressing exogenous protein in plant plastids such as
PT chloroplasts
XX Example 1; Page 17; 79pp; English.
XX This is the nucleotide sequence of maize rpoB gene primer rpoB#3.
CC The 5' nucleotide of the primer corresponds to nucleotide 21394
CC of the complementary strand of the maize plastid genome sequence.
CC The primer was designed to add a XhoI restriction site downstream
CC of an amplified rpoB fragment following PCR amplification. The PCR
CC product was cloned into vector pBSKS+ and used to generate
CC protecting RNA for use in in vitro capping experiments. The
CC invention provides isolated rpoB, atpB, clpP and 16S rDNA NEP and
CC PEP promoter elements (see AAY9569-99) useful for producing
CC exogenous proteins of interest in plant plastids.
XX Sequence 27 BP; 9 A; 5 C; 6 G; 7 T; 0 other;

Query Match 62.0%; Score 12.4; DB 20; Length 27;
Best Local Similarity 92.9%; Pred. No. 1.5e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ttgctagatgctag 16
Db 17 TTGCTAGATTCCTAG 4
|||||

RESULT 14
AAAA8985
ID AAA48985 standard; DNA; 27 BP.
XX AAA48985;
AC AAA48985;
XX 28-NOV-2000 (first entry)
XX Mutagenic PCR primer Bcl2-3 targeted to anti-apoptotic gene bcl-2.
XX Apoptosis; bcl-2; resistance; bacterial; viral; pathogen;
KW PCR primer; human; ss.
XX Homo sapiens.
OS
XX WO200026391-A2.
PN 11-MAY-2000.
XX 29-OCT-1999; 99WO-US25522.
XX 30-OCT-1998; 98US-0106321.
PR 09-JUN-1999; 99US-0138303.
XX (UYNE-) UNIV NEBRASKA-LINCOLN.
PA

```

```

XX Dickman MB;
PI WPI; 2000-365634/31.
XX Transgenic plants with improved resistance characteristics comprising
DR nucleic acids encoding apoptotic proteins -
XX Example 3; Page 58; 109pp; English.
XX The present invention relates to the use of apoptotic genes in the the
CC production of transgenic plants with improved resistance
CC characteristics. The present sequence is the mutagenic PCR primer Bcl2-3.
CC This primer was used with primer Bcl2-5 (AAA48984) to introduce a 5' NcoI
CC site and a 3' XbaI site in the human apoptotic gene bcl-2. Specifically
CC the primer introduced an Ala residue between the Met (position 1) and
CC His (position 2) residues of the native protein. The altered bcl-2
CC sequence was used in the creation of the final vector used to transform
CC plants. The improved resistance characteristics of the plants helps
CC protect against bacterial, viral and other pathogens. Resistance to
CC abiotic challenges may also be conferred.
XX Sequence 27 BP; 2 A; 5 C; 8 G; 12 T; 0 other;

Query Match 62.0%; Score 12.4; DB 21; Length 27;
Best Local Similarity 92.9%; Pred. No. 1.5e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ttgtctagatgct 14
Db 2 ttttctagatgct 15
|||||

RESULT 15
AAT59549
ID AAT59549 standard; DNA; 29 BP.
XX AAT59549;
AC AAT59549;
XX 18-NOV-1997 (first entry)
XX Reverse primer amplifies fragment encoding ODV-E66 residues 1-23.
XX N-terminal transport polypeptide; insect pest control; cell cycle;
KW Autographa californica multinucleocapsid nuclear polyhedrosis virus;
KW AcMNPV; occlusion derived virus; ODV; 66 kD envelope protein; ODV-E66;
KW membrane protein; intranuclear viral-induced microvesicle; GFP; URF-13;
KW beta-galactosidase; b-gal; green fluorescent protein; marker; primer;
KW polymerase chain reaction; PCR; amplify; ss.
XX Synthetic.
OS
XX WO9703198-A2.
PN 30-JAN-1997.
XX 05-JUL-1996; 96WO-US11320.
XX 03-JUL-1996; 96US-0000955.
PR 07-JUL-1995; 95US-0000955.
XX (TEXA ) UNIV TEXAS A & M SYSTEM.
XX Braunagel SC, Hong TC, Summers MD;
PI WPI; 1997-119053/11.
XX New isolated transport polypeptide from AcMNPV - can direct
PT localisation of desired proteins to occlusion derived virus
PT envelopes
XX Example 1; Page 39; 104pp; English.
PA

```

XX The sequences given in AAT59548-49 are primers which were used to  
CC amplify the sequence encoding residues 1-23 of Autographa californica  
CC multi-nucleocapsid nuclear polyhedrosis virus (ACMNPV) occlusion  
CC derived virus (ODV) 66 kb envelope protein (ODV-E66). ODV-E66 is an  
CC integral membrane protein of the ODV envelope. In an infected cell  
CC nucleus, ODV-E66 is present in the intranuclear viral-induced  
CC microvesicles and ODV envelope, providing evidence that the microvesicles  
CC function as an immediate precursor in the assembly of the ODV envelope.  
CC In infected cells, ODV-E66 localises to the ODV envelope, intranuclear  
CC microvesicles, membranes of the nuclear envelope and cytoplasmic  
CC membranes in close juxtaposition to the nuclear envelope. The amplified  
CC sequence encodes the N-terminal, target sequence domain of ODV-E66. The  
CC target sequence may be used to target beta-galactosidase (b-gal), green  
CC fluorescent protein (GFP) and URF-13 to the membranes of the nuclear  
CC envelope which are further transported into the intranuclear  
CC microvesicles and ODV envelope. The target protein can be used to target  
CC foreign proteins to the ODV envelope and intranuclear microvesicles. It  
CC can be used for insect pest control, for therapeutic applications, e.g.  
CC for location of a protein, peptide or derivative into the nuclear  
CC membrane which will alter the cell cycle and for diagnostics, e.g. to  
CC insert a marker specific for a disease or abnormality into a cell tissue  
CC manifesting such abnormalities.

XX SQ Sequence 29 BP; 5 A; 3 C; 6 G; 15 T; 0 other;

Query Match 61.0%; Score 12.2; DB 18; Length 29;  
Best Local Similarity 82.4%; Pred. No. 1.9e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 ttgctagatgctaggta 19  
||||| |||||  
Db 13 ttgctatttgataggta 29

Search completed: October 2, 2001, 16:18:38  
Job time: 15482 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:37 ; Search time 876.95 Seconds  
(without alignments)  
14.320 Million cell updates/sec

Title: US-09-757-100B-8

Perfect score: 20

Sequence: 1 ctatagctaggtatctgtc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	22	AAC65540 Human focal adhesi
2	15.8	79.0	27	21	AAZ93950 CaWV Calibrator pr
3	15	75.0	15	22	AAC65560 Human focal adhesi
4	15	75.0	20	22	AAC65541 Human focal adhesi
c 5	13.4	67.0	43	14	AAQ50213 c-fos position 339
6	13	65.0	38	13	AAQ28274 PCR primer 2 for a
7	12.8	64.0	24	18	AAT97310 Human Fas signal s
c 8	12.8	64.0	27	21	AAZ37684 PCR primer for 1-2
c 9	12.8	64.0	39	21	AAZ87799 Human tumour suppr
10	12.6	63.0	21	16	AAT00720 Multiple tumour su
11	12.6	63.0	21	16	AAQ99928 Human MTS1E1-beta

12	12.6	63.0	21	18	AAT72309 Human multiple tum
13	12.6	63.0	21	19	AAV53849 Nucleotide sequenc
14	12.6	63.0	21	19	AAV11268 Human MTS1E1-beta
15	12.6	63.0	21	20	AAV70613 Reverse PCR primer
16	12.6	63.0	21	21	AAA95847 Human MTS1E1-beta g
17	12.6	63.0	21	21	AAA11178 Human multiple tum
18	12.6	63.0	21	21	AAA39383 Human MTS1E1-beta
19	12.6	63.0	21	21	AAZ48786 PCR primer for hum
20	12.6	63.0	21	21	AAZ39979 PCR primer for hum
21	12.6	63.0	21	22	AAE58201 Reverse primer #1.
c 22	12.6	63.0	21	22	AAC88772 Human catenin-bind
23	12.6	63.0	21	22	AAC60498 Reverse primer use
24	12.6	63.0	22	22	AAC88771 Human catenin-bind
c 25	12.6	63.0	37	14	AAQ49810 bGH gene 3' antise
26	12.6	63.0	39	16	AAQ85851 MHC class I allele
27	12.6	63.0	39	16	AAQ85781 MHC CW3 forward am
28	12.6	63.0	39	16	AAQ85745 MHC CW3 CDNA forwa
29	12.6	63.0	39	20	AAZ31530 Probe EF-4 for HIV
c 30	12.6	63.0	39	20	AAZ11016 Probe EF-7 for HIV
c 31	12.4	62.0	33	21	AAZ72026 HIV-1 proviral DNA
c 32	12.4	62.0	33	21	AAZ88080 HIV packaging cons
c 33	12.4	62.0	33	21	AAZ56555 pHP-1dl 28 WT nucl
34	12.4	62.0	35	19	AAV34760 Human KDR primer 2
c 35	12.2	61.0	18	13	AAQ34171 Downstream sequenc
c 36	12.2	61.0	25	21	AAZ48824 PCR primer for Hum
37	12.2	61.0	36	21	AAZ99288 Nucleotide sequenc
38	12.2	61.0	36	21	AAZ99289 Nucleotide sequenc
39	12.2	61.0	42	18	AAT86866 plasmid pBS-HIV ol
40	12.2	61.0	42	20	AAV64882 HIV anti-viral oli
41	12	60.0	15	22	AAC65561 Human focal adhesi
42	12	60.0	17	21	AAZ03205 Hammerhead ribozym
43	12	60.0	17	21	AAZ03206 Primer for HIV RNA
c 44	12	60.0	24	19	AAV15580 PCR primer used to
45	12	60.0	25	22	AAZ24892

#### ALIGNMENTS

RESULT 1  
AAC65540  
ID AAC65540 standard; DNA; 20 BP.  
XX AC AAC65540;  
XX AC AAC65540;  
DT 12-FEB-2001 (first entry)  
XX Human focal adhesion kinase antisense sequence #6.  
XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.  
XX Homo sapiens.  
XX US6133031-A.  
XX 17-OCT-2000.  
XX 19-AUG-1999; 99US-0377310.  
XX 19-AUG-1999; 99US-0377310.  
XX (ISIS-) ISIS PHARM INC.  
XX Monia BP, Gaarde WA;  
XX WPI; 2001-006141/01.  
XX New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -  
XX

PS Claim 3; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
 CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
 CC is involved in integrin-mediated signal transduction, and is implicated  
 CC in cancer, particularly colon, breast and oral tumours, embryonic  
 CC development disorders, angiogenic disorders and wound healing. The  
 CC antisense sequences, including the one shown here, can be used in the  
 CC treatment of all of these.

XX Sequence 20 BP; 4 A; 4 C; 5 G; 7 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.18;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgtc 20

|||||

Db 1 ctatagctaggtatctgtc 20

RESULT 2

AAZ93950

ID AAZ93950 standard; DNA; 27 BP.

AC AAZ93950;

XX 29-AUG-2000 (first entry)

XX CamV Calibrator probe sequence used in detection method.

XX Detection; identification; PCR; polymerase chain reaction; probe;

KW primer; calibrator; diagnosis; pathogen; virus; bacteria; HHV-6;

KW HHV-7; HHV-8; human herpes virus; human immunodeficiency virus; HIV;

KW AIDS; acquired immune deficiency syndrome;

KW Mycobacterium tuberculosis; ss.

XX Synthetic.

OS WO200029613-A1.

PN 25-MAY-2000.

XX 17-NOV-1999; 99WO-EP08847.

XX 17-NOV-1998; 98IT-MI02491.

PR (SANR-) FOND CENT SAN RAFFAELE DEL MONTE TABOR.

PA Locatelli G, Lusso P, Mainati M, Salvatori F, Scarlatti G;

XX WPI; 2000-387819/33.

XX Quantitative detection of nucleic acids, based on the use of a  
 PT calibrator, suitable primers and probes, and a nucleic acid polymerase  
 PT with 5'-3' nuclease activity

XX Example 1; Page 13; 39pp; English.

XX A new method for the quantitative detection of nucleic acids in a  
 CC sample which uses a calibrator, suitable primers and probes, and  
 CC a nucleic acid polymerase with 5'-3' nuclease activity is described.  
 CC The method comprises adding a calibrator nucleotide to the sample,  
 CC the calibrator nucleotide having the same sequence as the target with  
 CC the exception of one or more regions which in the target hybridize  
 CC with a probe labeled with a reporter and quencher, or which hybridize  
 CC with the probe and with 2 or more primers. These regions have a  
 CC different, randomized nucleotide sequence and a similar Tm. Then  
 CC extracting the calibrator and target nucleotides from the sample.  
 CC Probes and primers are then added to the extracted sample/calibrator  
 CC mixture and PCR is performed. Reaction in the presence of the target  
 CC nucleic acid specific probe permits quantitation of the copy number

CC of the extracted target nucleic acid. The reaction in the presence of  
 CC the calibrator nucleotide permits quantitation of the calibrator copy  
 CC number. The reaction in the presence of both permits calculation of  
 CC the total number of target templates and calibrator, allowing  
 CC calculation of the percentage of calibrator recovery. The method may  
 CC be used for the diagnosis of viral and any other pathogenic agents in  
 CC body fluids, and to monitor safety and/or genetic composition of  
 CC waters, foods, and plant species used in the alimentary field. The  
 CC method is used to quantitatively detect the genomic nucleic acid of  
 CC human herpes virus (HHV)-6, HHV-7, HHV-8 and human immunodeficiency  
 CC virus (HIV). The method provides an enhanced sensitivity, accuracy,  
 CC and precision and a reduced measure viability. For specific probe,  
 CC primer and calibrator sequences used in the method, see GENESEQ  
 XX records AAZ93935-293965.

SQ Sequence 27 BP; 5 A; 7 C; 7 G; 8 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 27;

Best Local Similarity 89.5%; Pred. No. 28;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctatagctaggtatctgt 19

|||||

Db 5 ctatagctaggtatctgt 23

RESULT 3

AAZ93950

ID AAC65560 standard; DNA; 15 BP.

XX AAC65560;

XX 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #26.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer;

KW embryonic development disorder; angiogenic disorder; wound healing;

KW antisense; phosphorothioate; ss.

XX Homo sapiens.

XX US6133031-A.

XX 17-OCT-2000.

XX 19-AUG-1999; 99US-0377310.

XX 19-AUG-1999; 99US-0377310.

XX (ISIS-) ISIS PHARM INC.

XX Monia BP, Gaarde WA;

XX WPI; 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase

PT expression, especially useful for inhibiting retinal

PT neovascularization, or for diagnosing and treating e.g. colon cancer -

XX Example 2; Column 25; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
 CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
 CC is involved in integrin-mediated signal transduction, and is implicated  
 CC in cancer, particularly colon, breast and oral tumours, embryonic  
 CC development disorders, angiogenic disorders and wound healing. The  
 CC antisense sequences, including the one shown here, can be used in the  
 CC treatment of all of these.

XX Sequence 15 BP; 4 A; 2 C; 4 G; 5 T; 0 other;

Query Match 75.0%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 agatgctaggtatct 17  
 |||||  
 Db 1 agatgctaggtatct 15

RESULT 4  
 AAC65541  
 ID AAC65541 standard; DNA; 20 BP.  
 XX  
 AC AAC65541;  
 XX  
 DT 12-FEB-2001 (first entry)  
 XX  
 DE Human focal adhesion kinase antisense sequence #7.  
 XX  
 DE Human; focal adhesion kinase; FAK; signal transduction; cancer;  
 KW embryonic development disorder; angiogenic disorder; wound healing;  
 KW antisense; phosphorothioate; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6133031-A.  
 XX  
 PD 17-OCT-2000.  
 XX  
 PF 19-AUG-1999; 99US-0377310.  
 XX  
 PR 19-AUG-1999; 99US-0377310.  
 XX  
 PA (ISIS-) ISIS PHARM INC.  
 XX  
 PI Monla BP, Gaarde WA;  
 XX  
 DR WPI; 2001-006141/01.  
 XX  
 XX New antisense compounds for inhibiting focal adhesion kinase  
 PT expression, especially useful for inhibiting retinal  
 PT neovascularization, or for diagnosing and treating e.g. colon cancer -  
 XX  
 PS Claim 3; Column 23; 30pp; English.  
 XX  
 CC The present invention describes a number of phosphorothioate antisense  
 CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
 CC is involved in integrin-mediated signal transduction, and is implicated  
 CC in cancer, particularly colon, breast and oral tumours, embryonic  
 CC development disorders, angiogenic disorders and wound healing. The  
 CC antisense sequences, including the one shown here, can be used in the  
 CC treatment of all of these.  
 XX  
 SQ Sequence 20 BP; 4 A; 2 C; 5 G; 9 T; 0 other;

Query Match 75.0%; Score 15; DB 22; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctatagctaggtat 15  
 |||||  
 Db 6 ctatagctaggtat 20

RESULT 5  
 AAQ50213/c  
 ID AAQ50213 standard; DNA; 43 BP.  
 XX  
 AC AAQ50213;  
 XX  
 DT 06-MAY-1994 (first entry)

Query Match 75.0%; Score 13.4; DB 14; Length 43;  
 Best Local Similarity 93.3%; Pred. No. 5.1e+02;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gatgctaggtatctg 18  
 |||||  
 Db 19 GAAGCTAGGTATCTG 5

RESULT 6  
 AAQ28274  
 ID AAQ28274 standard; cDNA; 38 BP.  
 XX  
 AC AAQ28274;  
 XX  
 DT 15-FEB-1993 (first entry)  
 XX  
 DE PCR primer 2 for a novel type III RTK gene - the KDR gene.  
 KW Receptor tyrosine kinase; vascular endothelial cell growth factors;  
 KW cancer; tumour; diagnosing; monitoring; ss.

XX  
 DE C-fos position 3392-3434 INS mutagenic oligonucleotide.  
 XX  
 KW HIV-1; p17gag; inhibition; gag; M1; M2; M3; M4; silent mutation; mRNA;  
 KW vector; p17; point mutation; p17M1234; Hltat cells; gene replacement;  
 KW inhibitory/instability signal; INS; stability; utilisation; vaccine;  
 KW interferon; interleukin; fos proto-oncogene protein; growth factor;  
 KW env; attenuated; AIDS; AIDS-related disease; latent infection;  
 KW gene therapy; Human immunodeficiency virus type 1; p24;  
 KW long terminal repeat; LTR; ss.  
 XX

## Synthetic.

WO9320212-A.

14-OCT-1993.

29-MAR-1993; 93WO-US02908.

27-MAR-1992; 92US-0858747.

(USSH ) US SEC DEPT HEALTH.

Felber BK, Pavlakis GN;

WPI; 1993-336919/42.

Eliminating inhibitory-instability regions in mRNA to improve  
 stability and expression - by making multiple point mutations  
 within A-T rich regions

Example 3; Page 69; 117pp; English.

The sequences given in AAQ50202-70 are oligonucleotides which were  
 used to illustrate the method of the invention for the mutagenesis of  
 inhibitory/instability signals (INS). Mutation in an INS produced  
 using oligonucleotides such as these, increases the stability and/or  
 utilisation of mRNA without changing its protein coding capacity, or  
 if the sequence is changed, its function is maintained. Other genes  
 encoding such mRNA molecules include growth factor, interferon,  
 interleukin, fos proto-oncogene protein and HIV env and gag gene  
 proteins. Nucleic acid constructs in which INS function has been  
 impeded, can be used as vaccines, esp. against AIDS and AIDS-related  
 diseases by preventing HIV from establishing a latent infection, as  
 is possible using the INS, and thus escaping immune system  
 surveillance. The constructs may also be used in gene therapy for  
 gene replacement by homologous recombination with a target gene in  
 situ. See also AAQ50200-02.

Sequence 43 BP; 9 A; 11 C; 5 G; 18 T; 0 other;

Query Match 67.0%; Score 13.4; DB 14; Length 43;

Best Local Similarity 93.3%; Pred. NO. 5.1e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gatgctaggtatctg 18

Db 19 GAAGCTAGGTATCTG 5

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT modified_base 15
XX FT /*tag= a
XX FT /mod_base= I
XX PN W09214748-A.
XX PD 03-SEP-1992.
XX PF 20-FEB-1992; 92WO-US01300.
XX PR 22-FEB-1991; 91US-0657236.
XX PA (AMCY ) AMERICAN CYANAMID CO.
XX PI Carrion ME, Terman BI;
XX DR WPI; 1992-316117/38.
XX PT DNA encoding type III receptor tyrosine kinase - useful for
XX PT diagnosing the onset of cancer
XX PS Claim 16; Fig 2; 101pp; English.
XX CC This PCR primer is designed from a region of the kinase domain 5'
XX CC to the kinase insert domain of a consensus sequence of known type III
XX CC RTK gene catalytic domains. It is used with AAQ28273 to amplify human
XX CC endothelial cDNA producing 251 and 420 bp products. Sequencing of the
XX CC 251bp product revealed a novel sequence containing both primers but with
XX CC little homology to known tyrosine kinases.
XX SQ Sequence 38 BP; 5 A; 10 C; 8 G; 10 T; 5 other;

Query Match 65.0%; Score 13; DB 13; Length 38;
Best Local Similarity 68.4%; Pred. No. 8.1e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ctatagctagtagtctgt 19
Db 19 ctrgcgcagctctcgt 37

RESULT 7
ID AAT97310 standard; DNA; 24 BP.
XX AC AAT97310;
XX DT 06-MAR-1998 (first entry)
XX DE Human Fas signal sequence generating oligonucleotide Fas1.
XX KW human Fas ligand; hFasL; fusion protein; truncated; T lymphocyte;
XX KW glycosyl-phosphatidylinositol; GPI; treatment; allograft; xenograft;
XX KW Fas signal sequence; ss.
XX OS Synthetic.
XX OS Homo sapiens.
XX PN W09718307-A1.
XX PD 22-MAY-1997.
XX PF 15-NOV-1996; 96WO-EP05039.
XX PR 16-NOV-1995; 95GB-0023469.
XX PA (SANO ) SANDOZ LTD.
XX PI (SANO ) SANDOZ PAT.

```

```

PA (SANO ) SANDOZ-ERF.
XX PI Buehler T;
XX DR WPI; 1997-289283/26.
XX PT Human Fas Ligand fused to carboxy-terminal glycopospholipid -
XX PT useful for preventing or treating tissue or organ allograft or
XX PT xenograft rejection
XX PS Example 1; Page 11; 26pp; English.
XX CC This oligonucleotide Fas1 is used for the generation of a Fas signal
XX CC sequence for post translational modification of a novel human Fas ligand
XX CC (hFasL) fusion protein. This novel protein comprises of hFasL or a
XX CC truncated or functionally equivalent variant that retains the Fas
XX CC receptor binding and apoptosis inducing properties of hFasL. This is
XX CC linked to glycosyl-phosphatidylinositol (GPI) at its C-terminus. DNA
XX CC encoding hFasL amino acids, a linker sequence and a human Fas signal
XX CC GPI addition signal sequence derived from human CD16. The fusion protein
XX CC can be produced by culturing COS cells transformed by the expression
XX CC vector product. This novel fusion protein can incorporate its lipid tail
XX CC into cell membranes, e.g. endothelial cell membranes, and thus present
XX CC the FasL protein on the cell surface. This can bind to Fas receptor
XX CC present on other cells, particularly T lymphocytes, and thereby induce
XX CC their apoptosis. This is useful for preventing or treating tissue or
XX CC organ allograft or xenograft rejection. This provides a specific
XX CC treatment for activated T lymphocytes, i.e. only for T lymphocytes that
XX CC expresses the Fas antigen, which attack the transplanted tissue or organ.
XX SQ Sequence 24 BP; 3 A; 5 C; 9 G; 7 T; 0 other;

Query Match 64.0%; Score 12.8; DB 18; Length 24;
Best Local Similarity 87.5%; Pred. No. 9.9e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agatgctaggtatctg 18
Db 8 agatgctggggatctg 23

RESULT 8
ID AAA37684/c
XX AC AAA37684;
XX DT 24-OCT-2000 (first entry)
XX DE PCR primer for 1-2-rhamnosyl-transferase cDNA sequence.
XX KW 1-2-rhamnosyl-transferase; hesperidin conversion; orange peel; NHDC;
XX KW flavone-7-O-glucosidase-2'-O-rhamnosyl-transferase; flavanoid glycoside;
XX KW sweetener; neohesperidin dihydrochalcone; grapefruit; pomelo;
XX KW citrus fruit; PCR primer; ss.
XX OS Citrus sp.
XX PN WO200043490-A2.
XX PD 27-JUL-2000.
XX PF 20-JAN-2000; 2000WO-IL00038.
XX PR 22-JAN-1999; 99IL-0128193.
XX PA (VEDA ) VEDA RES & DEV CO LTD.
XX PA (ISRA ) ISRAEL MIN AGRIC.
XX PI Gressel J, Eyal Y, Fluhr R;
XX

```



DR WPI; 2000-499220/44.  
 XX Polynucleotide encoding a  
 PT flavone-7-O-glucosidase-2"-O-rhamnosyl-transferase, useful for  
 PT converting hesperidin from orange peels to the sweetener neohesperidin  
 PT dihydrochalcone (NHDC) -  
 XX  
 XX Example; Page 24; 48pp; English.  
 XX  
 XX This sequence represents a PCR primer for DNA encoding a  
 CC 1'-2-rhamnosyl-transferase. The invention relates  
 CC to a polynucleotide encoding a polypeptide with  
 CC flavone-7-O-glucosidase-2"-O-rhamnosyl-transferase catalytic activity  
 CC (such as 1-2-rhamnosyl-transferase). The enzyme is useful for converting  
 CC hesperidin from orange peels to the sweetener neohesperidin  
 CC dihydrochalcone (NHDC), and to provide genetically modified plants of the  
 CC Citrus genus including an antisense or sense (for co-suppression)  
 CC construct, or knockout integrated construct, to provide less bitter  
 CC grapefruits, pomeios and other citrus fruits containing flavanoid  
 CC glycosides.  
 XX  
 XX Sequence 27 BP; 11 A; 6 C; 5 G; 5 T; 0 other;  
 SQ

Query Match 64.0%; Score 12.8; DB 21; Length 27;  
 Best Local Similarity 87.5%; Pred. No. 1e+03;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctatagctaggtatc 16  
 || ||||| |||||  
 DB 27 CTTGATGCTTGGTATC 12

RESULT 9  
 AA287799/C  
 ID AA287799 standard; DNA; 39 BP.  
 XX  
 XX AA287799;  
 XX  
 XX 12-MAY-2000 (first entry)  
 XX  
 XX Human tumour suppressor TMPRSS2 cDNA sequencing primer 1B.  
 XX  
 XX Tumour suppressor gene; TMPRSS2; cancer; human; drug design;  
 KW gene therapy; protein therapy; PCR primer; ss.  
 KW Homo sapiens.  
 OS  
 XX  
 XX WO200000605-A1.  
 XX  
 XX 06-JAN-2000.  
 XX  
 XX 29-JUN-1999; 99WO-US14622.  
 XX  
 XX 29-JUN-1998; 98US-0091044.  
 XX  
 XX (MYRI-) MYRIAD GENETICS INC.  
 XX  
 XX Wong AKC, Tavtigian SV, Teng DHF;  
 PI  
 XX  
 XX WPI; 2000-170914/15.  
 DR  
 XX Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of  
 PT human cancer -  
 XX  
 XX Example 2; Page 53; 89pp; English.  
 XX  
 XX The invention provides a new tumour suppressor gene, designated TMPRSS2.  
 CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for  
 CC diagnosing and prognosing predisposition to cancer in humans. The  
 CC polypeptides may also be used in assays to screen for compounds with  
 CC anti-cancer or therapeutic properties. The polypeptides are also useful  
 CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides

CC may be used for gene therapy and protein therapy. Sequences AA287797-812  
 CC represent PCR primers for sequencing the TMPRSS2 cDNA.  
 XX  
 XX Sequence 39 BP; 9 A; 11 C; 7 G; 12 T; 0 other;  
 SQ

Query Match 64.0%; Score 12.8; DB 21; Length 39;  
 Best Local Similarity 87.5%; Pred. No. 1e+03;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 atgctaggtatctgtc 20  
 ||| ||||| |||||  
 DB 30 ATGATAGGTATCCGTC 15

RESULT 10  
 AA100720  
 ID AA100720 standard; cDNA; 21 BP.  
 XX  
 XX AA100720;  
 AC  
 XX  
 XX 08-MAY-1996 (first entry)  
 XX  
 XX Multiple tumour suppressor 1 exon.1 beta gene PCR reverse primer.  
 XX  
 XX Multiple tumour suppressor; MTS1Elbeta; cancer; diagnosis; assay;  
 KW predisposition; melanoma; leukaemia; lymphoma; prognosis;  
 KW pancreas; breast; thyroid; PCR reverse primer; exon 1; ss.  
 XX  
 XX Synthetic.  
 OS  
 XX  
 XX WO9525813-A1.  
 XX  
 XX 28-SEP-1995.  
 XX  
 XX 17-MAR-1995; 95WO-US03537.  
 XX  
 XX 01-JUN-1994; 94US-0251938.  
 PR 18-MAR-1994; 94US-0214582.  
 PR 18-MAR-1994; 94US-0215086.  
 PR 18-MAR-1994; 94US-0215087.  
 PR 14-APR-1994; 94US-0227369.  
 XX  
 XX (MYRI-) MYRIAD GENETICS INC.  
 PA (UTAH ) UNIV UTAH RES FOUND.  
 XX  
 XX Cannon-Albright LA, Kamb A, Skolnick MH;  
 PI  
 XX  
 XX WPI; 1995-344626/44.  
 XX  
 XX Detecting polymorphism associated with cancer pre:disposition - also  
 PT DNA, vectors and host cells e.g. for gene or protein replacement  
 PT therapy and drug screening  
 PT  
 XX  
 XX Example 10; Page 64; 148pp; English.  
 XX  
 XX An individual can be diagnosed as having a predisposition to cancer  
 CC by detecting an alteration in the wild type multiple tumour  
 CC suppressor (MTS) gene, using gene probes which hybridise to the  
 CC MTS1Elbeta gene ORF (amplified using the PCR primers AA100719-21).  
 CC The above assay can also be used in the diagnosis and prognosis of  
 CC melanoma, lymphoma, leukaemia and pancreas, breast and thyroid  
 CC cancers, etc..  
 XX  
 XX Sequence 21 BP; 5 A; 3 C; 7 G; 6 T; 0 other;  
 SQ

Query Match 63.0%; Score 12.6; DB 16; Length 21;  
 Best Local Similarity 78.9%; Pred. No. 1.2e+03;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctatagctaggtatctgt 19  
 ||||| ||| |||||

```

Db      3   ctaggagggaattatctgt 21

RESULT 11
AAQ99928
ID      AAQ99928 standard; cDNA; 21 BP.
AC      AAQ99928;
XX
DT      07-MAY-1996 (first entry)
XX
DE      Human MTS1E1-beta PCR amplification primer.
XX
KW      Multiple tumour suppressor; E1-alpha; diagnosis; cancer; leukaemia;
KW      astrocytoma; glioblastoma; Hodgkin's lymphoma; melanoma; glioma;
KW      gene therapy; chronic; ss.
XX
OS      Synthetic.
XX
PN      WO9525429-A1.
XX
PD      28-SEP-1995.
XX
PF      17-MAR-1995; 95WO-US03316.
XX
PR      01-JUN-1994; 94US-0251938.
PR      18-MAR-1994; 94US-0214581.
PR      18-MAR-1994; 94US-0214582.
PR      18-MAR-1994; 94US-0215086.
PR      14-APR-1994; 94US-0227369.
PR      18-MAR-1994; 94US-0215086.
PR      18-MAR-1994; 94US-0215087.
XX
PA      (MYRI-) MYRIAD GENETICS INC.
XX
PI      Kamb A;
XX
WPI: 1995-344401/44.
XX
PT      Wild-type multiple tumour suppressor (MTS) gene and mutant sequences
PT      - useful in diagnosis, prognosis and therapy of human cancer, e.g.
PT      melanoma or leukaemia
XX
PS      Example 10; Page 64; 156pp; English.
XX
CC      The cDNA sequences encoding several multiple tumour suppressor (MTS)
CC      polypeptides have been isolated and sequenced, using various
CC      sequencing and amplification primers such as the primer represented
CC      in this sequence. MTS polypeptide-encoding cDNAs and mutants of
CC      these are useful for the diagnosis or prognosis of human cancer.
CC      Germ-line mutations of MTS cDNAs can be used for diagnosing
CC      predisposition to melanoma, leukaemia, astrocytoma, glioblastoma,
CC      lymphoma, glioma, Hodgkin's lymphoma, CLL and cancers of the pancreas,
CC      thyroid, ovary, uterus, testis, kidney, stomach and rectum. The
CC      wild-type gene is useful for gene therapy and MTS polypeptides may also
CC      be used for protein replacement therapy. Also the polypeptides or cells
CC      contg. an altered MTS gene are useful for screening for potential
CC      cancer therapeutics.
XX
SQ      Sequence 21 BP; 5 A; 3 C; 7 G; 6 T; 0 other;

Query Match      63.0%; Score 12.6; DB 16; Length 21;
Best Local Similarity 78.9%; Pred. No. 1.2e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      1   ctagatgctaggatctgt 19
        ||||| || | |||||
Db      3   ctaggagggaattatctgt 21

RESULT 12
AAT72309
ID      AAT72309 standard; DNA; 21 BP.
AC      AAT72309;
XX
DT      10-SEP-1997 (first entry)
XX
DE      Human multiple tumour suppressor gene 1 Elbeta reverse primer.
XX
KW      Primer; polymerase chain reaction; PCR; amplification; Elbeta;
KW      human; multiple; tumour; suppressor; MTS1; cancer; diagnosis; ss.
XX
OS      Synthetic.
XX
PN      US5624819-A.
XX
PD      29-APR-1997.
XX
PF      18-MAR-1994; 94US-0214582.
XX
PR      07-JUN-1995; 95US-0474177.
PR      18-MAR-1994; 94US-0214582.
PR      18-MAR-1994; 94US-0215086.
PR      18-MAR-1994; 94US-0215087.
PR      14-APR-1994; 94US-0227369.
PR      01-JUN-1994; 94US-0251938.
PR      17-MAR-1995; 95WO-US03537.
XX
PA      (MYRI-) MYRIAD GENETICS INC.
PA      (UTAH ) UNIV UTAH RES FOUND.
XX
PI      Cannon-Albright LA, Kamb A, Skolnick MH;
XX
WPI: 1997-258217/23.
XX
PT      Human mutant multiple tumour suppressor gene sequences - for
PT      production of recombinant mutant polypeptide(s)
XX
PS      Example 10; Columns 87-88; 72pp; English.
XX
CC      The present sequence is primer for the PCR amplification of the
CC      human multiple tumour suppressor gene 1 Elbeta (MTS1Elbeta),
CC      useful in cancer diagnosis.
XX
SQ      Sequence 21 BP; 5 A; 3 C; 7 G; 6 T; 0 other;

Query Match      63.0%; Score 12.6; DB 18; Length 21;
Best Local Similarity 78.9%; Pred. No. 1.2e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      1   ctagatgctaggatctgt 19
        ||||| || | |||||
Db      3   ctaggagggaattatctgt 21

RESULT 13
AAV53849
ID      AAV53849 standard; DNA; 21 BP.
AC      AAV53849;
XX
DT      04-DEC-1998 (first entry)
XX
DE      Nucleotide sequence of PCR primer 13.
XX
KW      Multiple tumour suppressor; MTS; human; cancer; hybridisation;
KW      somatic mutation; gene therapy; PCR; primer; amplification; ss.
XX
OS      Synthetic.
XX
PN      US5801236-A.
XX
PD      01-SEP-1998.

```



XX Example 11; Column 48; 80pp; English.  
PS  
XX  
CC PCR primers AAV70612-14 were used to amplify nucleic acid encoding a  
CC human multiple tumour suppressor 1E1-beta (MTE1E1-beta) protein.  
CC Primers designed from the gene can be used to design primers to  
CC detect abnormalities i.e. polymorphisms which may predispose  
CC towards malignancies such as melanoma, leukaemia, astrocytoma,  
CC lymphoma, glioma, as well as tumours of e.g. the breast, thyroid,  
CC pancreas, uterus and kidneys.  
XX  
SQ Sequence 21 BP; 5 A; 3 C; 7 G; 6 T; 0 other;  
  
Query Match 63.0%; Score 12.6; DB 20; Length 21;  
Best Local Similarity 78.9%; Pred. No. 1.2e+03;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 ctgagtgctagggtatctgt 19  
      ||||| | | |||||  
Db 3 ctgagggcgattatctgt 21

Search completed: October 2, 2001, 16:18:38  
Job time: 15482 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:36 ; Search time 876.95 seconds  
(without alignments)  
14.320 Million cell updates/sec

Title: US-09-757-100B-7

Perfect score: 20

Sequence: 1 ttctcccttcgattattctt 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	22	Human focal adhesi
2	15.2	76.0	36	20	Human prostate-spe
3	15	75.0	15	22	Human focal adhesi
4	13.8	69.0	33	19	LM609 grafted anti
5	13.8	69.0	33	22	DNA encoding mutan
6	13.4	67.0	46	19	M13mp18 template s
7	13.2	66.0	29	21	Polymorphic fragme
8	12.8	64.0	17	16	MDV L1 CDNA primer
9	12.8	64.0	16	20	PCR primer for PGI
10	12.8	64.0	21	20	Human alpha-7 nico
11	12.6	63.0	20	20	PCR primer used to

12	12.6	63.0	20	20	AA336892	Human XLIIS gene fr
13	12.6	63.0	24	20	AA230696	A. oryzae 40S ribo
14	12.6	63.0	34	17	AA242899	ss circular oligo
15	12.6	63.0	41	19	AAV50562	Brassica sp. polym
16	12.6	63.0	41	19	AAV50563	Brassica sp. polym
17	12.4	62.0	25	21	AAAG68425	Bacteriophage 3A O
18	12.4	62.0	25	21	AAAG68418	Bacteriophage 96 O
19	12.4	62.0	25	21	AAAG68819	Bacteriophage 96 O
20	12.2	61.0	25	21	AAAG68819	Human polymorphic
21	12.2	61.0	21	19	AA225981	Human tub gene 5'
22	12.2	61.0	22	18	AA225981	Human tub gene 5'
23	12.2	61.0	22	21	AA225981	PCR primer #2 for
24	12.2	61.0	32	16	AA225405	Human gene signatu
25	12.2	61.0	32	21	AA225405	Primer SIRrev to c
26	12.2	61.0	48	21	AA225405	Oligonucleotide us
27	12.2	61.0	48	21	AA225405	Oligonucleotide us
28	12.2	61.0	21	22	AA225405	Human gene single
29	12.2	61.0	26	21	AA225405	Arabidopsis thalia
30	12.2	61.0	26	21	AA225405	I-SceI gene partia
31	12.2	61.0	27	21	AA225405	I-SceI gene partia
32	12.2	61.0	30	14	AA225405	Nucleic acid ligan
33	12.2	61.0	30	14	AA225405	B-cell mRNA ribozy
34	12.2	61.0	30	17	AA225405	NGF SELEX clone ha
35	12.2	61.0	30	19	AA225405	SELEX ligand hsnf
36	12.2	61.0	30	19	AA225405	Human native inter
37	12.2	61.0	30	20	AA225405	SELEX identified l
38	12.2	61.0	30	20	AA225405	RNA ligand sequenc
39	12.2	61.0	30	21	AA225405	Nerve growth facto
40	12.2	61.0	34	17	AA225405	ss circular oligo
41	12.2	61.0	39	20	AA225405	Tomato prosystemin
42	12.2	61.0	39	20	AA225405	Tomato prosystemin
43	12.2	61.0	40	16	AA225405	Elastase DNA nucle
44	12.2	61.0	40	20	AA225405	Human brx PCR prim
45	12.2	61.0	43	16	AA225405	PCR primer oligo D
46	12.2	61.0	45	18	AA225405	DNA encoding unive

#### ALIGNMENTS

RESULT 1  
AAC65539  
ID AAC65539 standard; DNA; 20 BP.  
XX  
AC AAC65539;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human focal adhesion kinase antisense sequence #5.  
DE  
KW Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphothioate; ss.  
XX  
OS Homo sapiens.  
XX  
PN US6133031-A.  
XX  
PD 17-OCT-2000.  
XX  
PF 19-AUG-1999; 99US-0377310.  
XX  
PR 19-AUG-1999; 99US-0377310.  
XX  
PA (ISIS-) ISIS PHARM INC.  
XX  
PI Monia BP, Gaarde WA;  
XX  
DR M13mp18 template s  
XX  
XX Polymorphic fragme  
XX MDV L1 CDNA primer  
XX PCR primer for PGI  
XX Human alpha-7 nico  
XX PCR primer used to

PS Claim 3; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense sequences to the human focal adhesion kinase (FAK) protein. This protein is involved in integrin-mediated signal transduction, and is implicated in cancer, particularly colon, breast and oral tumours, embryonic development disorders, angiogenic disorders and wound healing. The antisense sequences, including the one shown here, can be used in the treatment of all of these.

XX Sequence 20 BP; 1 A; 7 C; 1 G; 11 T; 0 other;

SQ

Query Match 100.0%; Score 20; DB 22; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 0.68;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
 ||||| ||||| ||||| |||||

Db 1 ttctcccttcggtattctt 20

RESULT 2

AAx04289  
 ID AAX04289 standard; DNA; 36 BP.

XX AC AAX04289;

XX 15-APR-1999 (first entry)

XX Human prostate-specific antigen linker regions of PAP-290.

XX Ricin-like toxin; cancer; viral infection; parasitic infection; KW linker; B chain; A chain; protease; fungal infection; malaria; KW leucocyte proliferation; cytomegalovirus; herpes; hepatitis; KW rhinovirus; laryngeotracheitis; poliomyelitis; varicella zoster; KW cystic fibrosis; multiple sclerosis; ds.

XX Homo sapiens.

OS Synthetic.

XX WO9849311-A2.

XX 05-NOV-1998.

PD 30-APR-1998; 98WO-CA00394.

PF 29-OCT-1997; 97US-0063715.

PR 30-APR-1997; 97US-0045148.

XX (DNOV-) DE NOVO ENZYME CORP.

PA Borgford T;

PI WPI; 1999-009431/01.

XX New nucleic acid encoding ricin-like toxin with an interchain linker cleaved by protease - is specific for diseased cells, useful for, PT e.g. killing selectively cancer or infected cells

XX Disclosure; Fig 44B; 352pp; English.

XX The present invention describes new purified and isolated nucleic acids CC (I) encoding: (i) the A and B chains of a ricin-like toxin (II); and CC (ii) a heterologous linker, joining the two chains and including a CC cleavage recognition site for a disease-specific protease (III). Also CC described are: (1) plasmids or baculovirus transfer vectors that contain CC of (I); and (2) recombinant protein (IV) consisting of the A and B chains CC of (II) joined by the specified linker. (IV), produced by expression of CC (I) in host cells, are used to inhibit or kill diseased cells that CC produce (III), particularly for treating cancers (e.g. leucocyte CC proliferation; cancer of ovary, pancreas, breast or prostate; glioma) or CC infections caused by fungi, parasites (e.g. malaria) or viruses (e.g.

CC cytomegalovirus (CMV), herpes, hepatitis, rhinovirus, laryngeotracheitis, CC poliomyelitis or varicella zoster), also cystic fibrosis and multiple CC sclerosis. Alternatively, (I) is used to express (IV) in vivo. (IV) is CC toxic specifically for (III)-expressing cells and does not depend for CC specificity on a cell-binding component. When used to treat virus- CC infected cells, transcytosis and cytotoxicity of (IV) are increased by CC retrograde translocation from endoplasmic reticulum to cytoplasm (which CC some viruses exploit to avoid immune detection), so selectivity and CC safety are further improved. (IV) are not toxic until chain A is CC released and this occurs only in target cells. The present sequence CC represents a nucleotide sequence from the present invention.

XX Sequence 36 BP; 5 A; 9 C; 3 G; 19 T; 0 other;

SQ

Query Match 76.0%; Score 15.2; DB 20; Length 36;  
 Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttctcccttcggtattctt 20  
 ||||| ||||| ||||| |||||

Db 14 ttctcttcggtattctt 33

RESULT 3

AAc65559  
 ID AAC65559 standard; DNA; 15 BP.

XX AC AAC65559;

XX 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #25.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer; KW embryonic development disorder; angiogenic disorder; wound healing; KW antisense; phosphorothioate; ss.

XX Homo sapiens.

OS US6133031-A.

PN 17-OCT-2000.

PD 19-AUG-1999; 99US-0377310.

PF 19-AUG-1999; 99US-0377310.

PR (ISIS-) ISIS PHARM INC.

PA Monia BP, Gaarde WA;

PI WPI; 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase expression, especially useful for inhibiting retinal PT neovascularization, or for diagnosing and treating e.g. colon cancer - PT Example 2; Column 25; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense CC sequences to the human focal adhesion kinase (FAK) protein. This protein CC is involved in integrin-mediated signal transduction, and is implicated CC in cancer, particularly colon, breast and oral tumours, embryonic CC development disorders, angiogenic disorders and wound healing. The CC antisense sequences, including the one shown here, can be used in the CC treatment of all of these.

XX Sequence 15 BP; 1 A; 6 C; 1 G; 7 T; 0 other;

SQ

Query Match 75.0%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 ctccctccgttatt 17  
|||||  
Db 1 ctccctccgttatt 15

RESULT 4  
AAV49869  
ID AAV49869 standard; DNA; 33 BP.  
XX  
AC AAV49869;  
XX  
DT 02-NOV-1998 (first entry)  
XX  
DE LM609 grafted antibody V-L region CDR2 DNA fragment #2.  
XX  
KW Vitaxin; antibody; variable region; heavy chain; light chain; integrin;  
KW LM609; inhibitor; integrin-mediated signal transduction; treatment;  
KW diagnosis; angiogenesis; restenosis; inflammation; diabetic retinopathy;  
KW neovascular glaucoma; cancer; psoriasis; rheumatoid arthritis; murine;  
KW macular degeneration; osteoporosis; primer; V-L region; CDR;  
KW complementarity determining region; ss.  
OS Mus sp.  
XX  
XX  
PN WO9833919-A2.  
XX  
PD 06-AUG-1998.  
XX  
PF 30-JAN-1998; 98WO-US01826.  
XX  
PR 30-JAN-1997; 97US-0791391.  
XX  
PA (IXSY-) IXSYS INC.  
XX  
PI Glaser SM, Huse WD;  
XX  
DR WPI; 1998-437472/37.  
DR P-PSDB; AAW76032.  
XX  
XX Humanised antibody, Vitaxin, that binds selectively to alphaVbeta3  
PT integrin - and related grafted antibodies based on murine monoclonal  
PT LM609, also related nucleic acid, used to treat, prevent or diagnose  
PT angiogenesis or restenosis  
XX  
PS Disclosure; Page 42; 129pp; English.  
XX  
CC AAV49844-V49877 are nucleotide fragments of the grafted monoclonal  
CC antibody LM609 heavy and light chain variable region. LM609 and the  
CC antibody vitaxin bind selectively to integrin alphavbeta3 and can be used  
CC to inhibit binding of alphavbeta3 to a ligand and thus block  
CC integrin-mediated signal transduction. This is useful in the treatment,  
CC prevention and diagnosis of alphavbeta3-mediated disease, specifically  
CC angiogenesis and restenosis (but also e.g. (non-)immune inflammation,  
CC diabetic retinopathy, neovascular glaucoma, cancer, psoriasis,  
CC rheumatoid arthritis, macular degeneration, osteoporosis etc.). The  
CC antibodies contain non-murine framework regions so are suitable for use  
CC in humans. Enhanced types of LM609 have affinity more than 90 times  
CC greater than that of parent the parent antibody.  
XX  
SQ Sequence 33 BP; 4 A; 13 C; 2 G; 14 T; 0 other;

Query Match 69.08; Score 13.8; DB 19; Length 33;  
Best Local Similarity 88.2%; Pred. No. 6.5e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 4 ctccctccgttattctt 20  
|||||  
Db 3 tctcatcgcgttattctt 19

RESULT 5  
AAF28225  
ID AAF28225 standard; DNA; 33 BP.  
XX  
AC AAF28225;  
XX  
DT 03-APR-2001 (first entry)  
XX  
DE DNA encoding mutant VL CDR2 peptide.  
XX  
KW LM609; grafted antibody; alphaVbeta3 integrin; angiogenesis;  
KW inflammatory; cancer; retina; restenosis; osteoporosis; ss.  
XX  
OS Unidentified.  
XX  
PN WO200078815-A1.  
XX  
PD 28-DEC-2000.  
XX  
PF 23-JUN-2000; 2000WO-US17454.  
XX  
PR 24-JUN-1999; 99US-0339922.  
XX  
PA (MOLE-) APPLIED MOLECULAR EVOLUTION.  
XX  
XX Huse WD, Wu H;  
XX  
XX WPI; 2001-050110/06.  
XX  
PT Enhanced LM609 grafted antibodies exhibiting selective binding affinity  
PT to alpha(V)beta(3) integrin, useful in the diagnosis and treatment of  
PT angiogenesis, inflammatory diseases e.g. psoriasis, cancers and  
PT osteoporosis -  
XX  
PS Disclosure; Page 42; 132pp; English.  
XX  
XX The present invention relates to enhanced LM609 grafted antibodies  
CC exhibiting selective binding affinity to alphavbeta3 integrin or  
CC their functional fragments. The antibodies or their functional  
CC fragments can be used in the diagnosis and treatment of  
CC alphavbeta3-mediated diseases such as angiogenesis, inflammatory  
CC diseases (such as psoriasis and chronic articular rheumatism),  
CC disorders associated with inappropriate or inopportune invasion of  
CC vessels (such as diabetic retinopathy, neovascular glaucoma and  
CC cancer disorders such as tumours and Kaposi's sarcoma), retinal  
CC diseases (such as macular degeneration), restenosis and  
CC osteoporosis.  
XX  
SQ Sequence 33 BP; 4 A; 13 C; 2 G; 14 T; 0 other;

Query Match 69.08; Score 13.8; DB 22; Length 33;  
Best Local Similarity 88.2%; Pred. No. 6.5e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 4 tccctccgcgttattctt 20  
|||||  
Db 3 tctcatcgcgttattctt 19

RESULT 6  
AAV27760  
ID AAV27760 standard; DNA; 46 BP.  
XX  
AC AAV27760;  
XX  
DT 24-SEP-1998 (first entry)  
XX  
DE M13mp18 template sequence.  
XX  
KW ss; M13mp18; sequencing; target specificity; PCR; amplification;  
KW differential primer extension.  
XX





XX O'Connell PH, Ohashi K, Schat KA;  
 XX WPI; 1995-255063/33.  
 XX Marek's disease virus protein and its nucleotide sequence - used to  
 PT protect chickens against Marek's disease and inhibit the establishment  
 PT of latency and tumour cell development  
 XX  
 XX Example 1; Page 11; 42pp; English.  
 XX  
 XX The primers given in AAQ97859-60 were used for the PCR amplification  
 CC of the upstream region of MDV L1 cDNA. The amplified fragment,  
 CC Q2L1, was cloned into pBluescript KS+ for sequencing.  
 XX  
 XX Sequence 17 BP; 9 A; 2 C; 5 G; 1 T; 0 other;

Query Match 64.0%; Score 12.8; DB 16; Length 17;  
 Best Local Similarity 87.5%; Pred. NO. 1.8e+03;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 cccctccgttattctt 20  
 ||||| ||||| ||  
 Db 17 CCTTCCTTTATTGTT 2

RESULT 9  
 AAZ01237/C  
 ID AAZ01237 standard; DNA; 18 BP.  
 XX  
 XX AAZ01237;  
 XX  
 XX 27-SEP-1999 (first entry)  
 XX  
 XX PCR primer for PGI biallelic markers 4-54-283 and 4-55-95.  
 XX  
 XX PGI gene; biallelic marker; PCR primer; PGI-related biallelic marker;  
 KW cancer; prostate cancer; diagnosis; therapy; prostate specific antigen;  
 KW PSA; human; ss.  
 XX  
 XX Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO9932644-A2.  
 XX  
 XX 01-JUL-1999.  
 XX  
 XX 22-DEC-1998; 98WO-IB02133.  
 XX  
 XX 09-SEP-1998; 98US-0099658.  
 XX  
 XX 22-DEC-1997; 97US-0996306.  
 XX  
 XX (GEST ) GENSET.  
 XX  
 XX Blumenfeld M, Bougueleret L, Chumakov I, Cohen D;  
 PI WPI; 1999-405178/34.  
 XX  
 XX Use of a prostate cancer associated gene and biallelic markers  
 PT derived from it  
 XX  
 XX Claim 4; Page 354; 385pp; English.

XX The invention relates to a mammalian PGI gene and protein, and a set of  
 CC PGI biallelic markers. The PGI polynucleotide and biallelic markers are  
 CC used in a hybridisation assay, a sequencing assay, or in an  
 CC allele-specific amplification assay for determining the identity of a  
 CC nucleotide at a PGI-related biallelic marker. The methods can be used to  
 CC detect and to assess the risk of developing cancer or prostate cancer.  
 CC Early-stage diagnosis of prostate cancer relies on prostate specific  
 CC antigen (PSA) dosage. However, the effectiveness of this is limited due  
 CC to its inability to discriminate between malignant and non-malignant

CC affections of the organ. A need exists for both a reliable diagnostic  
 CC procedure which would enable early-stage diagnosis, and for preventative  
 CC and curative treatments of the disease. The PGI gene can be used for  
 CC detection of prostate cancer, and the risk of developing it in the  
 XX future, and can also be used to determine therapies for the disease.

SQ Sequence 18 BP; 6 A; 1 C; 8 G; 3 T; 0 other;

Query Match 64.0%; Score 12.8; DB 20; Length 18;  
 Best Local Similarity 87.5%; Pred. NO. 1.8e+03;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 ctccctccgttatttc 18  
 ||||| || ||||| ||  
 Db 16 CTCCTCCTCAGTTATTC 1

RESULT 10  
 AAX56167/C  
 ID AAX56167 standard; DNA; 21 BP.  
 XX  
 XX AAX56167;  
 AC  
 XX  
 XX 15-JUL-1999 (first entry)  
 XX  
 XX Human alpha-7 nicotinic receptor PCR primer SEQ ID NO:14.

XX Human; alpha-7 nicotinic receptor; neuronal; hybridisation; probe;  
 KW alpha-7 neuronal nicotinic acetylcholine receptor; schizophrenia;  
 KW small cell lung carcinoma; breast cancer; nicotine-dependent illness;  
 KW epilepsy; juvenile myoclonic epilepsy; Prader-Willi syndrome;  
 KW Angelman's syndrome; PCR primer; ss.

XX Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO9920757-A2.  
 XX  
 XX 29-APR-1999.  
 XX  
 XX 15-OCT-1998; 98WO-US21762.  
 XX  
 XX 23-OCT-1997; 97US-0956518.  
 XX  
 XX (FREE/) FREEDMAN R.  
 XX (LEON/) LEONARD S.  
 XX  
 XX Freedman R, Leonard S;  
 PI WPI; 1999-288306/24.  
 XX  
 XX Human alpha-7 neuronal nicotinic acetylcholine receptor and related  
 PT polynucleotides  
 XX  
 XX Claim 15; Page 64; 104pp; English.

XX The present invention describes an isolated nucleotide sequence (I)  
 CC encoding at least a portion of the human alpha-7 neuronal nicotinic  
 CC acetylcholine receptor (alpha7-hnAChR). Also described are: (1) a  
 CC peptide encoded by (I); (2) a vector comprising (I); (3) a host cell  
 CC transformed with a vector of (2); (4) a polynucleotide comprising at  
 CC least 15 nucleotides which hybridises under stringent conditions to at  
 CC least a portion of (I); (5) a method for detection of a polynucleotide  
 CC encoding alpha 7-hnAChR in a biological sample; and (6) a method for  
 CC amplification of nucleic acid from a sample suspected of containing  
 CC nucleic acid encoding alpha 7-hnAChR. The primers and probes from the  
 CC present invention can be used on brain tissue and blood samples of  
 CC humans suspected of suffering from schizophrenia, small cell lung  
 CC carcinoma, breast cancer and nicotine-dependent illness. This is  
 CC particularly useful for diagnosis of schizophrenia. Other illnesses  
 CC that can be studied/diagnosed are epilepsy (e.g. juvenile myoclonic  
 CC epilepsy) and Prader-Willi and Angelman's syndromes.

```

XX SQ Sequence 21 BP; 9 A; 3 C; 8 G; 1 T; 0 other;

Query Match          64.0%; Score 12.8; DB 20; Length 21;
Best Local Similarity 87.5%; Pred. No. 1.9e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tctcccttcggtatt 17
   ||||| |||||
Db 18 TCTCCCTTGGCTTCTT 3

RESULT 11
AAZ06154
ID AAZ06154 standard; DNA; 20 BP.
XX AC AAZ06154;
XX DT 07-OCT-1999 (first entry)
XX OS Synthetic.
XX OS PCR primer used to amplify an ORF of Chlamydia trachomatis.
XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX KW paratrachoma; inclusion conjunctivitis; genital disease; perithelatitis;
XX KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis; PCR primer;
XX KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX OS Synthetic.
XX OS Chlamydia trachomatis.
XX PN WO9928475-A2.
XX PD 10-JUN-1999.
XX PF 27-NOV-1998; 98WO-IB01939.
XX PR 04-NOV-1998; 98US-0107077.
XX PR 28-NOV-1997; 97FR-0015041.
XX PR 17-DEC-1997; 97FR-0016034.
XX PA (GEST ) GENSET.
XX PI Griffais R;
XX WPI; 1999-371125/31.
XX Genome sequence of Chlamydia trachomatis
XX Disclosure; Page 1829; 1755pp; English.
XX PCR primers AAZ01426-Z06209 were used to amplify open reading frames
XX (ORFs) of the genome of Chlamydia trachomatis (see AAZ01425). These ORFs
XX encode polypeptides (see AAY36754-Y37949) which can be used as vaccines
XX against Chlamydia trachomatis. Antisense and ribozyme sequences
XX can also be used to control growth of the microorganism. Chlamydia
XX trachomatis is responsible for a large number of diseases, e.g. eye
XX diseases such as conventional trachoma, nonendemic trachoma,
XX paratrachoma, and inclusion conjunctivitis; genital diseases such as
XX nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
XX perithelatitis, bartholinitis; pneumopathy in breast feeding infants;
XX CC and venereal lymphogranulomatosis. The polypeptides of the
XX CC invention may be of use in treating these diseases.
XX SQ Sequence 20 BP; 2 A; 8 C; 2 G; 8 T; 0 other;

Query Match          63.0%; Score 12.6; DB 20; Length 20;
Best Local Similarity 78.9%; Pred. No. 2.3e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 tctcccttcggtattctt 20
   | |||| | ||||| | |

Db 2 tttcccatcgcgttatccct 20

RESULT 12
AAZ36892
ID AAX36892 standard; DNA; 20 BP.
XX AC AAX36892;
XX DT 14-JUL-1999 (first entry)
XX DE Human XLIS gene fragment PCR primer 6 F.
XX KW XLIS gene; human; detection; diagnosis; prenatal diagnosis; therapy;
XX KW lissencephaly; LIS; agyria-pachygyria; subcortical laminar heterotopia;
XX KW SCLH; cortical dysgenesis; cryptogenic epilepsy; neurological disorder;
XX KW neurodegenerative disease; Alzheimer's disease; X-linked disorder;
XX KW genetic counselling; PCR primer; ss.
XX OS Synthetic.
XX OS Homo sapiens.
XX PN EP918091-A1.
XX PD 26-MAY-1999.
XX PF 21-NOV-1997; 97EP-0402811.
XX PR 21-NOV-1997; 97EP-0402811.
XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Chelly J, Des Portes V, Kahn A, Pinard J;
XX WPI; 1999-290318/25.
XX New gene and its gene product expressed in the brain, useful for
XX diagnosing and treating disorders such as lissencephaly and
XX subcortical laminar heterotopia
XX Claim 9; Page 50; 71pp; English.
XX This sequence is a primer for the human XLIS gene of the invention.
XX The XLIS fragments may be used to detect abnormalities in the expression
XX of the XLIS gene transcripts or to compare their sequence with that of
XX the XLIS transcripts from patients for in vitro especially prenatal
XX diagnosis of lissencephaly (LIS) (or agyria-pachygyria), subcortical
XX laminar heterotopia (SCLH), cortical dysgenesis, cryptogenic epilepsies
XX or neurodegenerative diseases such as Alzheimer's disease. These
XX disorders mainly affect females as the XLIS gene is X-linked. The XLIS
XX fragments may also be used to administer to patients to prevent or treat
XX the above disorders and may be used as a tool in genetic counselling.
XX Oligonucleotides which bind to the fragments may be used to amplify the
XX XLIS gene from a sample for comparison to normal samples in the in vitro
XX diagnosis regime. This may also be performed by amplifying XLIS cDNA from
XX the mRNA in the sample. Antibodies to XLIS may be used to detect XLIS in
XX a biological sample or can be administered to patients to prevent or
XX treat the above disorders. They may also be used to purify XLIS from a
XX biological sample. XLIS may also be administered to patients to prevent
XX or treat the above neurological disorders. In addition XLIS may be used
XX as a marker of neuronal cells at an early stage of development; its
XX discovery increases understanding of both the neuronal movement which
XX leads to development of the cortical region of the brain and of the
XX pathogenesis of the group of neuronal disorders mentioned above.
XX SQ Sequence 20 BP; 1 A; 7 C; 0 G; 12 T; 0 other;

Query Match          63.0%; Score 12.6; DB 20; Length 20;
Best Local Similarity 78.9%; Pred. No. 2.3e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tttcccttcggtattctt 19

```

Db 2 ttatcccttcctttctct 20  
 || ||||| || |||

## RESULT 13

AAZ30696/c  
 ID AAZ30696 standard; DNA; 24 BP.

XX  
 AC AAZ30696;

XX  
 DT 15-FEB-2000 (first entry)

XX  
 DE A. oryzae 40S ribosome protein S28 gene promoter primer.

XX  
 KW Promoter; 40S ribosomal protein S28; genetic engineering; amplification;  
 KW heterologous protein; gene expression; PCR; primer; ss.

XX  
 OS Synthetic.

XX  
 OS Aspergillus oryzae.

XX  
 PN JP11276170-A.

XX  
 PD 12-OCT-1999.

XX  
 PF 31-MAR-1998; 98JP-0105712.

XX  
 PR 31-MAR-1998; 98JP-0105712.

XX  
 XX (AMANO ) AMANO PHARM KK.

PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.

XX  
 DR WPI; 1999-626935/54.

XX  
 XX A new promoter derived from an Aspergillus genus microbe - useful for  
 PT producing exotic proteins

PS Example 7; Page 5; 11pp; Japanese.

XX  
 CC This primer was used to PCR amplify the promoter sequence from the  
 CC 40S ribosomal protein S28 gene (AAZ30685) from Aspergillus oryzae. The  
 CC invention relates to novel gene promoters (AAZ30680-230685) isolated  
 CC from Aspergillus oryzae which can be used in genetic engineering to  
 CC express heterologous proteins in Aspergillus.

XX  
 SQ Sequence 24 BP; 9 A; 6 C; 8 G; 1 T; 0 other;

Query Match 63.0%; Score 12.6; DB 20; Length 24;

Best Local Similarity 78.9%; Pred. No. 2.4e+03;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 tctcccttcctttctctt 20

Db 19 TCCTCCCTTCGGTGTGCTT 1  
 ||||| ||| |||

## RESULT 14

AAZ42899

ID AAT42899 standard; DNA; 34 BP.

XX  
 AC AAT42899;

XX  
 DT 10-JUN-1997 (first entry)

XX  
 DE ss circular oligo targetted to -385 bcr breakpoint.

XX  
 KW single stranded; circular; target sequence; parallel; detection;  
 KW binding domain; anti-parallel; loop domain; complementarity; ss;  
 KW synthesis; regulation; drug delivery; biosynthesis; tumour cell.

XX  
 OS Synthetic.

XX  
 FH Key

Location/Qualifiers

FT misc\_feature 1  
 FT /\*tag= a  
 FT /note= "forms 5' > 3' bond with C34"  
 FT 1..12  
 FT /\*tag= b  
 FT /note= "forms parallel binding domain"  
 FT 18..29  
 FT /\*tag= c  
 FT /note= "forms anti-parallel binding domain"  
 FT 34  
 FT /\*tag= d  
 FT /note= "forms 3' > 5' bond with T1"

FT  
 XX WO9630384-A1.

PN 03-OCT-1996.

XX 21-MAR-1996; 96WO-US03757.

XX 30-MAR-1995; 95US-0413813.

XX (RESE ) RESEARCH CORP TECHNOLOGIES INC.

XX Kool ET;

XX WPI; 1996-455262/45.

XX Single stranded circular oligo:nucleotide comprising parallel and or  
 PT anti-parallel binding domain - used to regulate biosynthesis of  
 PT DNA, RNA or protein in targetted mammalian tumour cell in vivo

XX Example 11; Page 135; 195pp; English.

XX The sequences given in AAT42898-901 are single stranded (ss) circular  
 CC oligonucleotides and their targets, which are used in the inhibition  
 CC of the proliferation of myeloid leukaemia cells. These oligos are  
 CC specifically targetted to a region in the bcr3/abl2 gene 385  
 CC nucleotides 5' to the bcr/abl junction, abd towards the bcr2/abl2  
 CC junction. These ss circular oligonucleotides comprise a parallel  
 CC binding (P) domain, and/or an anti-parallel binding (AP) domain, and  
 CC at least 1 loop domain. The P and AP domains have sufficient  
 CC complementarity to bind detectably to 1 strand of a defined nucleic  
 CC acid target. The P domain is capable of binding in a parallel manner  
 CC to the target. The AP domain is capable of binding in an anti-parallel  
 CC manner to the target and the ends of the P and AP domains are separated  
 CC by the loop domains. The ss circular oligonucleotides can be used to  
 CC regulate the synthesis of DNA, RNA or protein (pref. by DNA replication,  
 CC DNA reverse transcription, RNA splicing, RNA polyadenylation, RNA  
 CC translocation or protein translocation) by binding a target sequence  
 CC in the template. They can also be used to deliver a drug to a specific  
 CC cell type by administering a drug covalently bound to them (i.e. to  
 CC regulate the biosynthesis of DNA, RNA or protein in a targetted  
 CC mammalian tumour cell in vivo, without substantially altering the  
 CC biosynthesis of the DNA). They can also be used to detect a target  
 CC nucleic acid by detecting an oligonucleotide-target complex. The  
 CC circular oligonucleotide can bind both single and double stranded  
 CC target nucleic acids, and has enhanced stability, compared to  
 CC linear forms.

XX SQ Sequence 34 BP; 1 A; 18 C; 1 G; 14 T; 0 other;

Query Match 63.0%; Score 12.6; DB 17; Length 34;

Best Local Similarity 78.9%; Pred. No. 2.5e+03;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttctcccttcctgtttctt 19

Db 7 ttctcccttcctgtttctt 25  
 ||||| ||| |||

RESULT 15

AAV50562

ID AAV50562 standard; DNA; 41 BP.  
XX AC AAV50562;  
XX DT 21-DEC-1998 (first entry)  
XX DE Brassica sp. polymorphic marker N2/10B8/N3-2A DNA.  
XX KW Polymorphic marker; allele-specific; primer; probe; amplification;  
KW hybridisation; plant; hybrid certification; genetic contribution;  
KW progeny; back-cross; hybrid; ancestry; ss.  
XX OS Brassica sp.  
XX FH Key Location/Qualifiers  
FT variation 21  
FT /\*tag= a  
FT /replace= "c"  
FT /note= "polymorphism"  
XX PN W09824796-A1.  
XX PD 11-JUN-1998.  
XX PF 01-DEC-1997; 97WO-US21782.  
XX PR 07-MAR-1997; 97US-0813507.  
XX PR 02-DEC-1996; 96US-0032069.  
XX PA (AFFY-) AFFYMETRIX INC.  
XX PI Landry BS, Lemieux B, Murigneux A, Sapolsky RJ;  
XX WPI; 1998-333252/29.  
XX PT Brassica species allele-specific oligonucleotide probes and primers  
PT - useful for plant breeding  
XX PS Claim 1; Page 32; 65pp; English.  
XX CC This DNA sequence is a region of a Brassica napus or Brassica oleracea  
CC genome which contains a polymorphic marker. This sequence can be used  
CC in the construction of allele-specific primers and probes for  
CC amplification or hybridisation, e.g. to determine common or disparate  
CC ancestry between 2 or more plants, to monitor the genetic contribution  
CC of an ancestral plant, to trace the progeny of proprietary plants, in  
CC certification of a hybrid plant or to identify the progeny of a  
CC back-crossed plant with an ancestral plant.  
XX SQ Sequence 41 BP; 8 A; 11 C; 3 G; 19 T; 0 other;  
  
Query Match 63.0%; Score 12.6; DB 19; Length 41;  
Best Local Similarity 78.9%; Pred. No. 2.5e+03;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 ttctcccttcggtattct 19  
      ||||| | | | | | | | |  
Db 12 ttctccagttcttattct 30

Search completed: October 2, 2001, 16:18:37  
Job time: 15481 sec



;  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-26

Query Match 75.0%; Score 15; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aactgcagaagggcac 17  
| | | | | | | | | | | | | | |  
Db 1 aactgcagaagggcac 15

RESULT 3  
US-08-288-405A-3/c  
; Sequence 3, Application US/08288405A  
; Patent No. 5559009  
; GENERAL INFORMATION:  
; APPLICANT: Chandy, Kanianthara G.  
; APPLICANT: Kalman, Katalin  
; APPLICANT: Chandy, Grisha  
; APPLICANT: Gutman, George A.  
; TITLE OF INVENTION: A No. 5559009el Voltage-Gated Potassium Channel  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert,  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/288,405A  
; FILING DATE: 10-AUG-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/207,431  
; FILING DATE: 04-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dregler, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-59844-1/WH/D  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: both  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14..25  
US-08-288-405A-3

Query Match 69.0%; Score 13.8; DB 1; Length 25;  
Best Local Similarity 88.2%; Pred. No. 3e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaaactgcagaagggcac 17  
| | | | | | | | | | | | | | |  
Db 17 GGAAGTGCAGAAGGGAC 1

RESULT 4  
US-08-935-312-15/c  
; Sequence 15, Application US/08935312  
; Patent No. 6207455  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, Lung-Ji  
; TITLE OF INVENTION: LENTIVIRAL VECTORS  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.  
; STREET: 624 Ninth Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/935,312  
; FILING DATE: 22-SEP-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: CHANG-112  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 42 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "DNA"  
US-08-935-312-15

Query Match 68.0%; Score 13.6; DB 4; Length 42;  
Best Local Similarity 80.0%; Pred. No. 4e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaaactgcagaagggcac 20  
| | | | | | | | | | | | | | |  
Db 42 GAAAGAGCAGAAGACAGTGA 23

RESULT 5  
US-08-700-186-12/c  
; Sequence 12, Application US/08700186  
; Patent No. 5780286  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick  
; APPLICANT: Vockley, Joseph  
; TITLE OF INVENTION: ARGINASE II  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-2799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/700,186  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hap, William T  
REGISTRATION NUMBER: 34,344  
REFERENCE/DOCKET NUMBER: ATG50004-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5219  
TELEFAX: 610-270-5090  
TELEX:

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-700-186-12

Query Match 67.0%; Score 13.4; DB 1; Length 43;  
Best Local Similarity 93.3%; Pred. No. 5e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaagcact 18  
|||||  
DB 16 ACTGCAGAGGCAAT 2

## RESULT 6

US-08-914-981-12/c  
Sequence 12, Application US/08914981  
Patent No. 5912159  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick  
APPLICANT: Vockley, Joseph  
TITLE OF INVENTION: ARGINASE II  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/914,981  
FILING DATE: 20-AUGUST-1997  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/700,186  
FILING DATE: 20-AUGUST-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: ATG50004-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700

TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-914-981-12

Query Match 67.0%; Score 13.4; DB 2; Length 43;  
Best Local Similarity 93.3%; Pred. No. 5e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaagcact 18  
|||||  
DB 16 ACTGCAGAGGCAAT 2

## RESULT 7

US-09-116-115-12/c  
Sequence 12, Application US/091161115D  
Patent No. 6054308  
GENERAL INFORMATION:  
APPLICANT: VOCKLEY, JOSEPH G.  
APPLICANT: DILLON, PATRICK J.  
TITLE OF INVENTION: ARGINASE II  
FILE REFERENCE: ATG-50004-3/D1  
CURRENT APPLICATION NUMBER: US/09/116,115D  
CURRENT FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: US 08/914,981  
EARLIER FILING DATE: 1997-08-20  
EARLIER APPLICATION NUMBER: US 08/700,186  
EARLIER FILING DATE: 1996-08-20  
EARLIER APPLICATION NUMBER: US 60/013,395  
EARLIER FILING DATE: 1996-03-14  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 43  
TYPE: DNA  
ORGANISM: HOMO SAPIENS  
US-09-116-115-12

Query Match 67.0%; Score 13.4; DB 3; Length 43;  
Best Local Similarity 93.3%; Pred. No. 5e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 actgcagaagcact 18  
|||||  
DB 16 ACTGCAGAGGCAAT 2

## RESULT 8

US-07-825-959-13  
Sequence 13, Application US/07825959  
Patent No. 5372929  
GENERAL INFORMATION:  
APPLICANT: Cimino, George C.  
APPLICANT: Lin, Lilly  
TITLE OF INVENTION: METHOD FOR MEASURING THE INACTIVATION OF  
TITLE OF INVENTION: PATHOGENS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Limbach & Limbach  
STREET: 2001 Ferry Building

CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/825,959  
FILING DATE: 19920127  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Weseman, James C  
REGISTRATION NUMBER: 30,507  
REFERENCE/DOCKET NUMBER: HRI-02200  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 433-4150  
TELEFAX: (415) 433-8716  
TELEX: 278356  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FRAGMENT TYPE: internal  
US-07-825-959-13

Query Match 66.0%; Score 13.2; DB 1; Length 27;  
Best Local Similarity 83.3%; Pred. No. 5,7e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aactgcagaagcactga 20  
|||||  
Db 2 AACTGCAGAAGGCTAGGA 19

RESULT 9  
US-08-131-324-13  
; Sequence 13, Application US/08131324  
; Patent No. 5565320  
; GENERAL INFORMATION:  
; APPLICANT: Cimino, George C.  
; APPLICANT: Lin, Lily  
; TITLE OF INVENTION: COMPOUNDS FOR THE PHOTODECONTAMINATION  
; TITLE OF INVENTION: OF PATHOGENS IN BLOOD  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steritech, Inc.  
; STREET: 2525 Stanwell Drive  
; CITY: Concord  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94520  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/131,324  
; FILING DATE: 28-JUN-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/825,959  
; FILING DATE: 27-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837

; REFERENCE/DOCKET NUMBER: STER-1100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 603-9071  
; TELEFAX: (510) 603-9099  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 27 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-131-324-13

Query Match 66.0%; Score 13.2; DB 1; Length 27;  
Best Local Similarity 83.3%; Pred. No. 5,7e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aactgcagaagcactga 20  
|||||  
Db 2 AACTGCAGAAGGCTAGGA 19

RESULT 10  
US-08-377-495-1  
; Sequence 1, Application US/08377495  
; Patent No. 5631137  
; GENERAL INFORMATION:  
; APPLICANT: Martin, Mark T.  
; APPLICANT: Smith, Rodger G.  
; APPLICANT: Darsley, Michael J.  
; APPLICANT: Simpson, David  
; APPLICANT: Blackburn, Gary F.  
; TITLE OF INVENTION: REACTION-BASED SCREENING FOR EXPRESSION  
; TITLE OF INVENTION: OF AND CONCENTRATION OF CATALYTIC MOIETIES  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/377,495  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/250,934  
; FILING DATE:  
; APPLICATION NUMBER: US/08/101,274  
; FILING DATE:  
; APPLICATION NUMBER: US/07/841,648  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Evans, Barry  
; REGISTRATION NUMBER: 22,802  
; REFERENCE/DOCKET NUMBER: 370068-3580  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)



US-08-377-495-1

Query Match 64.0%; Score 12.8; DB 1; Length 22;  
Best Local Similarity 87.5%; Pred. No. 8.6e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaaactgcagaagcca 16  
||||||| |||  
Db 5 GAAACTGCAGGAGTCA 20

RESULT 11

US-08-748-415-1  
; Sequence 1, Application US/08748415  
; Patent No. 5891648  
; GENERAL INFORMATION:  
; APPLICANT: Martin, Mark T.  
; APPLICANT: Smith, Rodger G.  
; APPLICANT: Darsley, Michael J.  
; APPLICANT: Simpson, David  
; APPLICANT: Blackburn, Gary F.  
; TITLE OF INVENTION: REACTION-BASED SCREENING FOR EXPRESSION  
; TITLE OF INVENTION: OF AND CONCENTRATION OF CATALYTIC MOIETIES  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/748,415  
FILING DATE: 13-NOV-1996  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/377,495  
FILING DATE:  
APPLICATION NUMBER: US/08/250,934  
FILING DATE:  
APPLICATION NUMBER: US/08/101,274  
FILING DATE:  
APPLICATION NUMBER: US/07/841,648  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Barry  
REGISTRATION NUMBER: 22,802  
REFERENCE/DOCKET NUMBER: 370068-3580  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-748-415-1

Query Match 64.0%; Score 12.8; DB 2; Length 22;  
Best Local Similarity 87.5%; Pred. No. 8.6e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaaactgcagaagcca 16  
||||||| |||

Db 5 GAAACTGCAGGAGTCA 20

RESULT 12  
US-08-747-654-1  
; Sequence 1, Application US/08747654  
; Patent No. 6121007  
; GENERAL INFORMATION:

APPLICANT: Martin, Mark T.  
APPLICANT: Smith, Rodger G.  
APPLICANT: Darsley, Michael J.  
APPLICANT: Simpson, David  
APPLICANT: Blackburn, Gary F.  
TITLE OF INVENTION: REACTION-BASED SCREENING FOR EXPRESSION  
TITLE OF INVENTION: OF AND CONCENTRATION OF CATALYTIC MOIETIES  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/747,654  
FILING DATE: 13-NOV-1996  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/377,495  
FILING DATE:  
APPLICATION NUMBER: US/08/250,934  
FILING DATE:  
APPLICATION NUMBER: US/08/101,274  
FILING DATE:  
APPLICATION NUMBER: US/07/841,648  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Barry  
REGISTRATION NUMBER: 22,802  
REFERENCE/DOCKET NUMBER: 370068-3580  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-747-654-1

Query Match 64.0%; Score 12.8; DB 3; Length 22;  
Best Local Similarity 87.5%; Pred. No. 8.6e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaaactgcagaagcca 16  
||||||| |||  
Db 5 GAAACTGCAGGAGTCA 20

RESULT 13

US-09-235-353-1  
; Sequence 1, Application US/09235353  
; Patent No. 6177270  
; GENERAL INFORMATION:  
; APPLICANT: Martin, Mark T.

APPLICANT: Smith, Rodger G.  
APPLICANT: Darsley, Michael J.  
APPLICANT: Simpson, David  
APPLICANT: Blackburn, Gary F.  
TITLE OF INVENTION: REACTION-BASED SCREENING FOR EXPRESSION  
OF AND CONCENTRATION OF CATALYTIC MOIETIES  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/235,353  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/748,415  
FILING DATE: 13-NOV-1996  
APPLICATION NUMBER: US/08/377,495  
FILING DATE:  
APPLICATION NUMBER: US/08/250,934  
FILING DATE:  
APPLICATION NUMBER: US/08/101,274  
FILING DATE:  
APPLICATION NUMBER: US/07/841,648  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Barry  
REGISTRATION NUMBER: 22,802  
REFERENCE/DOCKET NUMBER: 370068-3580  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-235-353-1

Query Match 64.0%; Score 12.8; DB 4; Length 22;  
Best Local Similarity 87.5%; Pred. No. 8.6e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaaactgcagaaggca 16  
||| ||||| |||  
Db 5 GAAACTGCAGGAGTCA 20

RESULT 14  
US-09-232-477-1  
Sequence 1, Application US/09232477  
Patent No. 6228846  
GENERAL INFORMATION:  
APPLICANT: AUDONNET, Jean-Christophe  
APPLICANT: BOUCHARDON, Annabelle  
APPLICANT: RIVIERE, Michel  
TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST CANINE  
TITLE OF INVENTION: PATHOLOGIES, IN PARTICULAR RESPIRATORY AND DIGESTIVE  
TITLE OF INVENTION: PATHOLOGIES  
FILE REFERENCE: 454313-2240  
CURRENT APPLICATION NUMBER: US/09/232,477

CURRENT FILING DATE: 1999-01-15  
EARLIER APPLICATION NUMBER: PCT/FR97/01316  
EARLIER FILING DATE: 1997-07-15  
EARLIER APPLICATION NUMBER: 96/09401  
EARLIER FILING DATE: 1996-07-19  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 35  
TYPE: DNA  
ORGANISM: Borrelia burgdorferi  
US-09-232-477-1

Query Match 64.0%; Score 12.8; DB 4; Length 35;  
Best Local Similarity 87.5%; Pred. No. 9.2e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aaactgcagaaggcac 17  
||||| ||||| |||  
Db 2 aaactgcagaatgctc 17

RESULT 15  
US-08-105-483-148  
Sequence 148, Application US/08105483  
Patent No. 5494807  
GENERAL INFORMATION:  
APPLICANT: Paolletti, Enzo  
TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
TITLE OF INVENTION: STRAIN  
NUMBER OF SEQUENCES: 462  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
ADDRESSEE: c/o William S. Frommer  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/105,483  
FILING DATE: 12-AUG-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/847,951  
FILING DATE: 06-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2400  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 148:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-105-483-148

Query Match 63.0%; Score 12.6; DB 1; Length 21;  
Best Local Similarity 78.9%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaaactgcagaaggcactg 19

Db 3 GAAAGAGCAGAGACAGTG 21

Search completed: October 2, 2001, 16:03:46  
Job time: 14590 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:03:44 ; Search time 417.38 Seconds  
(without alignments)  
9.071 Million cell updates/sec

Title: US-09-757-100B-4  
Perfect score: 20  
Sequence: 1 ggcgcgtgaagcgaaggca 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 460742

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-09-377-310-4
2	15	75.0	15	3	US-09-377-310-24
3	13.2	66.0	21	1	US-08-250-849-19
4	13.2	66.0	21	1	US-08-434-474-19
5	12.8	64.0	20	4	US-09-180-437-51
6	12.8	64.0	20	4	US-09-180-437-52
7	12.8	64.0	20	4	US-09-180-437-53
8	12.2	61.0	17	3	US-08-598-099-6
9	12.2	61.0	38	1	US-08-750-077-4
10	12.2	61.0	40	3	US-08-841-267-15
11	12.2	61.0	50	2	US-08-472-171-46
12	12.2	61.0	50	2	US-08-894-526-46
13	12.2	61.0	50	3	US-09-013-047-46
14	12.2	61.0	50	3	US-09-374-597-46
15	12	60.0	20	1	US-07-678-448A-4
16	12	60.0	20	4	US-09-560-594-41
17	12	60.0	27	3	US-09-006-597-16
18	11.8	59.0	15	4	US-09-180-437-217
19	11.8	59.0	20	4	US-09-180-437-50
20	11.6	58.0	27	1	US-08-624-545-32
21	11.6	58.0	43	1	US-08-428-733A-46
22	11.4	57.0	20	4	US-09-517-584A-22
23	11.4	57.0	20	4	US-09-180-437-48
24	11.4	57.0	20	4	US-09-180-437-49
25	11.2	56.0	17	3	US-08-598-099-5
26	11.2	56.0	20	3	US-09-280-799-185
27	11.2	56.0	25	5	PCT-US92-02977-4

28	11.2	56.0	25	5	PCT-US95-03032-7	Sequence 7, Appli
c 29	11.2	56.0	38	1	US-08-015-770B-72	Sequence 72, Appl
c 30	11.2	56.0	30	1	US-08-399-580B-11	Sequence 11, Appl
31	11.2	56.0	33	1	US-08-438-639-16	Sequence 16, Appl
32	11.2	56.0	33	1	US-07-813-338A-16	Sequence 16, Appl
33	11.2	56.0	33	3	US-08-441-971-91	Sequence 91, Appl
34	11.2	56.0	33	4	US-08-221-653-91	Sequence 91, Appl
35	11.2	56.0	33	4	US-08-442-144A-91	Sequence 91, Appl
36	11.2	56.0	37	2	US-08-570-155-8	Sequence 8, Appli
37	11.2	56.0	37	5	PCT-US95-02861-8	Sequence 8, Appli
c 38	11.2	56.0	38	1	US-07-841-662-21	Sequence 21, Appl
c 39	11.2	56.0	38	1	US-08-209-797-21	Sequence 21, Appl
c 40	11.2	56.0	38	1	US-08-669-685-21	Sequence 21, Appl
c 41	11.2	56.0	38	3	US-09-103-486-21	Sequence 21, Appl
c 42	11.2	56.0	38	5	PCT-US93-01557-21	Sequence 21, Appl
43	11.2	56.0	41	1	US-07-951-715A-87	Sequence 87, Appl
44	11.2	56.0	41	2	US-08-459-448A-87	Sequence 87, Appl
45	11.2	56.0	41	3	US-08-459-448A-87	Sequence 87, Appl

ALIGNMENTS

RESULT 1  
US-09-377-310-4  
; Sequence 4, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence

US-09-377-310-4  
Query Match 100.0%; Score 20; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaaggca 20  
|||||  
Db 1 ggcgcgtgaagcgaaggca 20

RESULT 2  
US-09-377-310-24  
; Sequence 24, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

US-09-377-310-4  
Query Match 100.0%; Score 20; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaaggca 20  
|||||  
Db 1 ggcgcgtgaagcgaaggca 20

RESULT 2  
US-09-377-310-24  
; Sequence 24, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

US-09-377-310-4  
Query Match 100.0%; Score 20; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaaggca 20  
|||||  
Db 1 ggcgcgtgaagcgaaggca 20

RESULT 2  
US-09-377-310-24  
; Sequence 24, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

US-09-377-310-4  
Query Match 100.0%; Score 20; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaaggca 20  
|||||  
Db 1 ggcgcgtgaagcgaaggca 20

; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-24

Query Match 75.0%; Score 15; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cgcctgaagcgaag 17  
|||||  
Db 1 cgcctgaagcgaag 15

RESULT 3  
US-08-250-849-19  
; Sequence 19, Application US/08250849  
; Patent No. 5567583  
; GENERAL INFORMATION:  
; APPLICANT: Chang-Ning J. Wang and Kai-  
; APPLICANT: Yuan Wu  
; TITLE OF INVENTION: METHOD FOR DETECTING A TARGET  
; TITLE OF INVENTION: NUCLEIC ACID  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/250,849  
; FILING DATE: 05/26/94  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/808,463  
; FILING DATE: December 16, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Y. Rocky Tsao  
; REGISTRATION NUMBER: 34,053  
; REFERENCE/DOCKET NUMBER: 06498/002001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TOPLOGY: linear  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-250-849-19

Query Match 66.0%; Score 13.2; DB 1; Length 21;  
Best Local Similarity 83.3%; Pred. No. 4e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cgcctgaagcgaagc 19  
|||||  
Db 1 GCGGCGTGAAGCGCGC 18

RESULT 4  
US-08-434-474-19  
; Sequence 19, Application US/08434474  
; Patent No. 5712386

; GENERAL INFORMATION:  
; APPLICANT: Wang et al.  
; TITLE OF INVENTION: METHOD FOR DETECTING A TARGET  
; TITLE OF INVENTION: NUCLEIC ACID  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/434,474  
; FILING DATE: 05/04/95  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/250,849  
; FILING DATE: 05/26/94  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Y. Rocky Tsao  
; REGISTRATION NUMBER: 34,053  
; REFERENCE/DOCKET NUMBER: 06498/002002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-434-474-19

Query Match 66.0%; Score 13.2; DB 1; Length 21;  
Best Local Similarity 83.3%; Pred. No. 4e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cgcctgaagcgaagc 19  
|||||  
Db 1 GCGGCGTGAAGCGCGC 18

RESULT 5  
US-09-180-437-51  
; Sequence 51, Application US/09180437  
; Patent No. 6251873  
; GENERAL INFORMATION:  
; APPLICANT: FUKUSAKO, Shioji  
; APPLICANT: MORISAWA, Yoshifumi  
; APPLICANT: KUSUYAMA, Takeshi  
; TITLE OF INVENTION: Antisense Compounds to CD14  
; FILE REFERENCE: 1110-209P  
; CURRENT APPLICATION NUMBER: US/09/180,437  
; CURRENT FILING DATE: 1998-11-06  
; EARLIER APPLICATION NUMBER: PCT/JP98/00953  
; EARLIER FILING DATE: 1998-03-09  
; EARLIER APPLICATION NUMBER: 09-053518 JAPAN  
; EARLIER FILING DATE: 1997-03-07  
; NUMBER OF SEQ ID NOS: 289  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 51  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: other nucleic

; OTHER INFORMATION: acid, synthetic DNA  
US-09-180-437-51

Query Match 64.0%; Score 12.8; DB 4; Length 20;  
Best Local Similarity 87.5%; Pred. No. 6.1e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 ccgtgaagcgaagca 20  
|| ||||| |||||  
Db 1 ccctgaagccaagca 16

## RESULT 6

US-09-180-437-52  
; Sequence 52, Application US/09180437  
; Patent No. 6251873  
; GENERAL INFORMATION:  
; APPLICANT: FUKUSAKO, Shioji  
; APPLICANT: MORISAWA, Yoshifumi  
; APPLICANT: KUSUYAMA, Takeshi  
; TITLE OF INVENTION: Antisense Compounds to CD14  
; FILE REFERENCE: 1110-209P  
; CURRENT APPLICATION NUMBER: US/09/180,437  
; CURRENT FILING DATE: 1998-11-06  
; EARLIER APPLICATION NUMBER: PCT/JP98/00953  
; EARLIER FILING DATE: 1998-03-09  
; EARLIER APPLICATION NUMBER: 09-053518 JAPAN  
; EARLIER FILING DATE: 1997-03-07  
; NUMBER OF SEQ ID NOS: 289  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 52  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:other nucleic  
; OTHER INFORMATION: acid, synthetic DNA  
US-09-180-437-52

Query Match 64.0%; Score 12.8; DB 4; Length 20;  
Best Local Similarity 87.5%; Pred. No. 6.1e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 ccgtgaagcgaagca 20  
|| ||||| |||||  
Db 1 ccctgaagccaagca 16

## RESULT 7

US-09-180-437-53  
; Sequence 53, Application US/09180437  
; Patent No. 6251873  
; GENERAL INFORMATION:  
; APPLICANT: FUKUSAKO, Shioji  
; APPLICANT: MORISAWA, Yoshifumi  
; APPLICANT: KUSUYAMA, Takeshi  
; TITLE OF INVENTION: Antisense Compounds to CD14  
; FILE REFERENCE: 1110-209P  
; CURRENT APPLICATION NUMBER: US/09/180,437  
; CURRENT FILING DATE: 1998-11-06  
; EARLIER APPLICATION NUMBER: PCT/JP98/00953  
; EARLIER FILING DATE: 1998-03-09  
; EARLIER APPLICATION NUMBER: 09-053518 JAPAN  
; EARLIER FILING DATE: 1997-03-07  
; NUMBER OF SEQ ID NOS: 289  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 53  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:other nucleic  
; OTHER INFORMATION: acid, synthetic DNA  
US-09-180-437-53

Query Match 64.0%; Score 12.8; DB 4; Length 20;  
Best Local Similarity 87.5%; Pred. No. 6.1e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 ccgtgaagcgaagca 20  
|| ||||| |||||  
Db 4 ccctgaagccaagca 19

## RESULT 8

US-08-998-099-6/C  
; Sequence 6, Application US/08998099A  
; Patent No. 6103890  
; GENERAL INFORMATION:  
; APPLICANT: JARVIS, THALE  
; APPLICANT: MCSWIGGEN, JAMES A.  
; APPLICANT: STINCHCOMB, DAN T.  
; TITLE OF INVENTION: ENZYMTIC NUCLEIC ACID TREATMENT OF DISEASES  
; FILE REFERENCE: 231/175  
; CURRENT APPLICATION NUMBER: US/08/998,099A  
; CURRENT FILING DATE: 1997-12-24  
; EARLIER APPLICATION NUMBER: 60/037,658  
; EARLIER FILING DATE: 1997-01-23  
; EARLIER APPLICATION NUMBER: 08/373,124  
; EARLIER FILING DATE: 1995-01-13  
; EARLIER APPLICATION NUMBER: 08/245,466  
; EARLIER FILING DATE: 1994-05-18  
; NUMBER OF SEQ ID NOS: 375  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 17  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-08-998-099-6

Query Match 61.0%; Score 12.2; DB 3; Length 17;  
Best Local Similarity 82.4%; Pred. No. 1.2e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ggcgcgcgtgaagcgaag 17  
||||| ||||| ||  
Db 17 GGCCTGTGAAGCAGAG 1

## RESULT 9

US-08-750-077-4  
; Sequence 4, Application US/08750077  
; Patent No. 5783423  
; GENERAL INFORMATION:  
; APPLICANT: Wood, Patricia C  
; APPLICANT: Quirk, Alan V  
; TITLE OF INVENTION: Yeast Strains  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Centeon LLC  
; STREET: 1020 First Avenue  
; CITY: King of Prussia  
; COUNTRY: USA  
; ZIP: PA 19406-1310  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/750,077

```
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411356.0
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/01317
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Moore, Steven J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 878 4073
; TELEFAX: (610) 878 4221
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..38
; OTHER INFORMATION: /note="Oligonucleotide for PCR
; OTHER INFORMATION: amplification of the 3' end of the Hsp150 gene."
US-08-750-077-4

Query Match 61.0%; Score 12.2; DB 1; Length 38;
Best Local Similarity 82.4%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 gccgtgaagcgaagca 20
Db 1 GCCGTGACGAGGAA 17

RESULT 10
US-08-841-267-15
; Sequence 15, Application US/08841267C
; Patent No. 6007990
; GENERAL INFORMATION:
; APPLICANT: Levine, Robert A.
; APPLICANT: Wardlaw, Stephen C.
; TITLE OF INVENTION: Detection and Quantification of One or
; TITLE OF INVENTION: More Nucleotide Sequence Target Analytes in a Sample Using
; Patent No. 6007990
; TITLE OF INVENTION: Spatially Localized Target Analyte Replication
; FILE REFERENCE: UFB-001
; CURRENT APPLICATION NUMBER: US/08/841.267C
; CURRENT FILING DATE: 1997-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 40
; TYPE: DNA
; ORGANISM: virus HBV
US-08-841-267-15

Query Match 61.0%; Score 12.2; DB 3; Length 40;
Best Local Similarity 82.4%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggcgcgcgtgaagcgaag 17
Db 16 gacacggtgaagcgaag 32

RESULT 11
US-08-472-171-46
; Sequence 46, Application US/08472171
```

```
; Patent No. 5932714
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, Suite 701
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,171
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/393,334
; FILING DATE: 23-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-171-46

Query Match 61.0%; Score 12.2; DB 2; Length 50;
Best Local Similarity 82.4%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggcgcgcgtgaagcgaag 17
Db 23 GCGCGCGTGACGCGCAAG 39

RESULT 12
US-08-894-526-46
; Sequence 46, Application US/08894526
; Patent No. 5942418
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM
; TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
```



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,526  
FILING DATE: 01-DEC-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-724 MTS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-894-526-46

Query Match 61.0%; Score 12.2; DB 2; Length 50;  
Best Local Similarity 82.4%; Pred. No. 1.2e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggccgcgtgaagcgaag 17  
||||| | | | |

Db 23 GGCGCGGTGACGCGAAG 39

## RESULT 13

US-09-013-047-46  
Sequence 46, Application US/09013047  
Patent No. 5998168  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin H.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression Of Gene Products From  
Genetically Manipulated Strains Of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 330 University Avenue, 6th Floor  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/013,047  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/472,171  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/393,334  
FILING DATE: 23-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-507 MTS:vg  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-013-047-46

Query Match 61.0%; Score 12.2; DB 2; Length 50;  
Best Local Similarity 82.4%; Pred. No. 1.2e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggccgcgtgaagcgaag 17  
||||| | | | |

Db 23 GGCGCGGTGACGCGAAG 39

## RESULT 14

US-09-374-597-46  
Sequence 46, Application US/09374597  
Patent No. 6140082  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin H.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression Of Gene Products From  
Genetically Manipulated Strains Of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 330 University Avenue, 6th Floor  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/374,597  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/393,334  
FILING DATE: FEBRUARY 23, 1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-374-597-46

Query Match 61.0%; Score 12.2; DB 3; Length 50;  
Best Local Similarity 82.4%; Pred. No. 1.2e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggccgcgtgaagcgaag 17  
|||||  
Db 23 GGCGCGTGACGCGCAAG 39

## RESULT 15

US-07-678-448A-4/c  
; Sequence 4, Application US/07678448A  
; Patent No. 5521301  
; GENERAL INFORMATION:  
; APPLICANT: R. Bruce Wallace  
; APPLICANT: Luis Ugozzoli  
; TITLE OF INVENTION: Genotyping of Multiple Allele  
; TITLE OF INVENTION: Systems  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: City of Hope  
; STREET: 1500 East Duarte Road  
; CITY: Duarte  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 91010-0269  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3M Double Density 5 1/4" diskette  
; COMPUTER: Wang PC  
; OPERATING SYSTEM: MS-DOS (R) Version 3.30  
; SOFTWARE: Microsoft (R)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/678,448A  
; FILING DATE: 19910401  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/283,142  
; FILING DATE: 12 December, 1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Irons, Edward S.  
; REGISTRATION NUMBER: 16,541  
; REFERENCE/DOCKET NUMBER: No. 5521301e  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 785-6938  
; TELEFAX: (202) 785-5351  
; TELEX: 440087 LM WSH  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; MOLECULE TYPE: No. 5521301 Applicable  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: No. 5521301 Applicable  
; ORIGINAL SOURCE: Synthetically Prepared  
; IMMEDIATE SOURCE: Synthetically Prepared  
; POSITION IN GENOME: No. 5521301e  
; FEATURE: No. 5521301e  
; PUBLICATION INFORMATION: No. 5521301e  
US-07-678-448A-4

Query Match 60.0%; Score 12; DB 1; Length 20;  
Best Local Similarity 75.0%; Pred. NO. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggccgcgtgaagcgaagca 20  
|||||  
Db 20 GGCGCCTACAGCGCTGGCA 1

Search completed: October 2, 2001, 16:03:45  
Job time: 14589 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:33 ; Search time 876.95 seconds

(without alignments)  
14.320 Million cell updates/sec

Title: US-09-757-100B-4

Perfect score: 20

Sequence: 1 ggcgcgtgaagcgaaggca 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601:\*

1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT:\*

2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT:\*

3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT:\*

4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT:\*

5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT:\*

6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT:\*

7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT:\*

8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT:\*

9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT:\*

10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT:\*

11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT:\*

12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT:\*

13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT:\*

14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT:\*

15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT:\*

16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT:\*

17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT:\*

18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT:\*

19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT:\*

20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT:\*

21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:\*

22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	22	AAC65536 Human focal adhesi
2	15	75.0	15	22	AAC65556 Human focal adhesi
3	14.2	71.0	40	18	AAT87270 IL-4 2'F/NH2 RNA 1
4	14.2	71.0	40	18	AAT87201 Interleukin-4 2'F
5	13.2	66.0	21	17	AAT09720 Human AML-1 limit
6	13.2	66.0	21	18	AAT45484 Human leukaemia br
7	13.2	66.0	21	19	AAV09361 Limiting primer us
8	12.8	64.0	21	22	AAE95308 Human gene single
9	12.6	63.0	33	21	AAZ43694 Mass spectrometric
10	12.6	63.0	36	19	AAV36074 Oligonucleotide SC
11	12.6	63.0	50	21	AAA96608 Nucleotide sequenc

c	12	12.2	61.0	17	19	AAV95266 Human c-fos target
c	13	12.2	61.0	19	21	AAA84390 Cyclin D3 ribozyme
c	14	12.2	61.0	20	22	AAAF74126 Primer #60. Homo
c	15	12.2	61.0	22	22	AAAF69782 Human IL4Ra1pha ge
c	16	12.2	61.0	29	21	AAA04651 Polymorphic fragme
c	17	12.2	61.0	31	13	AAQ25255 NANBH PCR primer p
c	18	12.2	61.0	38	17	AAT07272 Primer for HBV pX
c	19	12.2	61.0	40	19	AAV08311 Primer for HBV pX
c	20	12.2	61.0	50	10	AAAN91945 Complementaty stra
c	21	12	60.0	20	17	AAAT30041 Allele-specific PC
c	22	12	60.0	21	22	AAAF23962 Bacillus lichenifo
c	23	12	60.0	27	19	AAAV07603 Nucleotide sequenc
c	24	12	60.0	36	21	AAA35716 permutein linker e
c	25	12	60.0	38	13	AAQ25438 probe for alkaline
c	26	12	60.0	38	17	AAT34327 Alkaline protease
c	27	12	60.0	50	21	AAAT77403 Human clone cg4492
c	28	11.8	59.0	20	20	AAZ00590 Human GPC4 exon 2
c	29	11.8	59.0	30	22	AAAF73589 HGF nucleic acid 1
c	30	11.8	59.0	42	20	AAAT21515 Integrin alpha 6 s
c	31	11.6	58.0	22	21	AAAD01060 Oligo #5 for site-
c	32	11.6	58.0	24	21	AAAC58525 Human PRO1434 (UNQ
c	33	11.6	58.0	24	21	AAAS1273 Forward primer for
c	34	11.6	58.0	25	21	AAAS30072 PCR primer 34387.t
c	35	11.6	58.0	25	22	AAAF60392 PRO240 probe #2.
c	36	11.6	58.0	27	21	AAAG60931 Phanerochaete sord
c	37	11.6	58.0	34	21	AAZ43239 PCR primer for C.
c	38	11.6	58.0	35	14	AAQ51295 D. nodosus proteas
c	39	11.6	58.0	35	20	AAAS1843 Porphyromonas ging
c	40	11.6	58.0	36	18	AAAT69182 Oligonucleotide Kb
c	41	11.6	58.0	38	20	AAAT78995 HIF-1alpha gene am
c	42	11.6	58.0	43	16	AAT02527 TnI gene forward p
c	43	11.6	58.0	43	19	AAV32360 Human troponin I f
c	44	11.6	58.0	43	20	AAZ25907 Human troponin I g
c	45	11.6	58.0	48	18	AAT97348 Mouse CGRP recepto

ALIGNMENTS

RESULT 1

AAC65536

ID AAC65536 standard; DNA; 20 BP.

XX

AC AAC65536;

XX

DT 12-FEB-2001 (first entry)

XX

DE Human focal adhesion kinase antisense sequence #2.

XX

KW Human; focal adhesion kinase; FAK; signal transduction; cancer;

KW embryonic development disorder; angiogenic disorder; wound healing;

KW antisense; phosphothioate; ss.

XX

OS Homo sapiens.

XX

PN US6133031-A.

XX

PD 17-OCT-2000.

XX

PF 19-AUG-1999; 99US-0377310.

XX

PR 19-AUG-1999; 99US-0377310.

XX

PA (ISIS-) ISIS PHARM INC.

PI Monia BP, Gaarde WA;

XX

DR WPI; 2001-006141/01.

XX

PT New antisense compounds for inhibiting focal adhesion kinase

PT expression, especially useful for inhibiting retinal

XX

PT neovascularization, or for diagnosing and treating e.g. colon cancer -

PS Claim 3; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, embryonic  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

XX Sequence 20 BP; 5 A; 5 C; 9 G; 1 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcgtgaagcgaagca 20

|||||

Db 1 ggcgcgtgaagcgaagca 20

RESULT 2

AAC65556

ID AAC65556 standard; DNA; 15 BP.

AC AAC65556;

DT 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #22.

DE Human; focal adhesion kinase; FAK; signal transduction; cancer;

KW embryonic development disorder; angiogenic disorder; wound healing;

KW antisense; phosphorothioate; ss.

XX Homo sapiens.

OS US6133031-A.

PN 17-OCT-2000.

XX 19-AUG-1999; 99US-0377310.

PF 19-AUG-1999; 99US-0377310.

PR (ISIS-) ISIS PHARM INC.

PA Monia BP, Gaarde WA;

PI WPI; 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase

PT expression, especially useful for inhibiting retinal

PT neovascularization, or for diagnosing and treating e.g. colon cancer -

XX Example 2; Column 25; 30pp; English.

PS The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, embryonic  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

XX Sequence 15 BP; 4 A; 4 C; 6 G; 1 T; 0 other;

Query Match

Best Local Similarity 75.0%; Score 15; DB 22; Length 15;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cgccgtgaagcgaag 17

|||||

Db 1 cgccgtgaagcgaag 15

RESULT 3

AAT87270

ID AAT87270 standard; RNA; 40 BP.

XX AAT87270;

XX 16-NOV-1997 (first entry)

XX IL-4 2'/NH2 RNA ligand, nitrocellulose filter binding clone #21.

XX high affinity ligand; cytokine; interferon-gamma; IFN-gamma; RANTES;

KW interleukin-4; IL-4; tumour necrosis factor-alpha; TNF-alpha;

KW Systematic Evolution of Ligands by Exponential enrichment; SELEX;

KW diagnosis; inflammatory response; septic shock; arthritis;

KW graft-vs-host reaction; ss.

XX Synthetic.

XX Key Location/Qualifiers

FH modified\_base 1..40

FT /\*tag= a

FT /note= "all U's are 2'-NH2 modified and all C's are

FT 2'-F modified"

XX WO9640717-A1.

PN 19-DEC-1996.

XX 04-JUN-1996; 96WO-US09537.

XX 07-JUN-1995; 95US-0481710.

PR 07-JUN-1995; 95US-0477527.

XX (NEXS-) NEXSTAR PHARM INC.

XX Gold L, Jayasena S, Pagratis N, Tasset D;

XX WPI; 1997-087049/08.

XX Identification of nucleic acid ligands that bind cytokine(s) - by

PT partitioning the ligands from a nucleic acid mixture, using SELEX

PT techniques

XX Claim 20; Page 69; 175pp; English.

XX AAT87178-276 are interleukin (IL)-4 RNA ligands obtained after 17 rounds  
CC of SELEX (Systematic Evolution of Ligands by Exponential enrichment).

CC The sequences were analysed for conserved sequences and aligned by this

CC criterion. The 2'F sequences fell into a single group with 12 orphan

CC sequences. 2'NH2 sequences fell into 2 distinct groups. Group 1 were

CC shown to bind to IL-4 and the other group were shown to bind to

CC nitrocellulose filters and were identified by the presence of a direct

CC repeat of the sequence GGAGG. A single orphan 2'NH2 sequence was also

CC found. The 2'F/NH2 sequences fell into 3 groups, of which one group

CC bound to nitrocellulose filters. Nucleic acid (NA) ligands to a cytokine

CC can be identified using SELEX. A candidate mixture of NA's are contacted

CC with a cytokine where the NA's having an increased affinity to the

CC cytokine relative to the candidate mixture may be partitioned from the

CC rest of the mixture. The NA's with increased affinity are amplified to

CC yield a mixture of NA's enriched for the NA sequences and relatively

CC higher affinity and selectivity for binding to the cytokine. The NA

CC ligands are useful in diagnostic and therapeutic applications especially

CC to prevent or treat diseases or medical conditions in human patients,

CC e.g. associated with excessive cytokine production such as inflammatory

CC responses mediated by IFN-gamma or interleukin-4, septic shock,

CC arthritis or graft-vs-host reactions mediated by tumour necrosis

CC factor-alpha.

XX

SQ Sequence 40 BP; 12 A; 5 C; 18 G; 5 U; 0 other;

Query Match 71.0%; Score 14.2; DB 18; Length 40;  
Best Local Similarity 78.9%; Pred. No. 4.1e+02;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 gcgcgtgaagcgaagca 20  
|||||:|||||  
DB 5 gcgccaagaagcaagga 23

## RESULT 4

AAT87201  
ID AAT87201 standard; RNA; 40 BP.

XX AC AAT87201;

XX DT 08-NOV-1997 (first entry)

XX DE Interleukin-4 2'F RNA ligand, orphan clone #21.

XX KW high affinity ligand; cytokine; interferon-gamma; IFN-gamma; RANTES;  
KW Interleukin-4; IL-4; tumour necrosis factor-alpha; TNF-alpha;  
KW Systematic Evolution of Ligands by Exponential enrichment; SELEX;  
KW diagnosis; inflammatory response; septic shock; arthritis;  
KW graft-vs-host reaction; ss.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT modified\_base 1..40

XX FT /\*tag= a  
XX FT /note= "all pyrimidines are 2'-F"

XX PN W09640717-A1.

XX PD 19-DEC-1996.

XX PD 04-JUN-1996; 96WO-US09537.

XX PR 07-JUN-1995; 95US-0481710.

XX PR 07-JUN-1995; 95US-0477527.

XX PA (NEXS-) NEXSTAR PHARM INC.

XX PI Gold L, Jayasena S, Pagratis N, Tasset D;

XX DR WPI; 1997-087049/08.

XX PT Identification of nucleic acid ligands that bind cytokine(s) - by  
XX PT partitioning the ligands from a nucleic acid mixture, using SELEX  
XX PT techniques

XX PS Claim 20; Page 65; 175pp; English.

XX AAT87178-276 are interleukin (IL)-4 RNA ligands obtained after 17 rounds  
XX of SELEX (Systematic Evolution of Ligands by Exponential enrichment).  
XX The sequences were analysed for conserved sequences and aligned by this  
XX criterion. The 2'F sequences fell into a single group with 12 orphan  
XX sequences. 2'NH2 sequences fell into 2 distinct groups. Group 1 were  
XX shown to bind to IL-4 and the other group were shown to bind to  
XX nitrocellulose filters and were identified by the presence of a direct  
XX repeat of the sequence GGAGG. A single orphan 2'NH2 sequence was also  
XX found. The 2'F/NH2 sequences fell into 3 groups, of which one group  
XX bound to nitrocellulose filters. Nucleic acid (NA) ligands to a cytokine  
XX can be identified using SELEX. A candidate mixture of NA's are contacted  
XX with a cytokine where the NA's having an increased affinity to the  
XX cytokine relative to the candidate mixture may be partitioned from the  
XX rest of the mixture. The NA's with increased affinity are amplified to  
XX yield a mixture of NA's enriched for the NA sequences and relatively  
XX higher affinity and selectivity for binding to the cytokine. The NA  
XX ligands are useful in diagnostic and therapeutic applications especially

CC to prevent or treat diseases or medical conditions in human patients,  
CC e.g. associated with excessive cytokine production such as inflammatory  
CC responses mediated by IFN-gamma or interleukin-4, septic shock,  
CC arthritis or graft-vs-host reactions mediated by tumour necrosis  
CC factor-alpha.

SQ Sequence 40 BP; 12 A; 5 C; 18 G; 5 U; 0 other;

Query Match 71.0%; Score 14.2; DB 18; Length 40;

Best Local Similarity 78.9%; Pred. No. 4.1e+02;

Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 gcgcgtgaagcgaagca 20

|||||:|||||

DB 5 gcgccaagaagcaagga 23

## RESULT 5

AAT09720  
ID AAT09720 standard; DNA; 21 BP.

XX AC AAT09720;

XX DT 27-JUN-1996 (first entry)

XX DE Human AML-1 limiting primer, for asymmetric amplification.

XX KW Polymerase chain reaction; amplification; non-specific priming;  
KW blocking oligonucleotide; donor; acceptor; fluorophore;  
KW energy transfer; ligation; AML-1; acute myeloid leukaemia;  
KW breakpoint related sequence; ss.

XX OS Synthetic.

XX PN W09532306-A1.

XX PD 30-NOV-1995.

XX PD 23-MAY-1994; 94WO-US05767.

XX PR 23-MAY-1994; 94WO-US05767.

XX PA (BIOT-) BIOTRONICS CORP.

XX PI Wang CJ, Wu K;

XX DR WPI; 1996-020598/02.

XX PT Detecting target nucleic acid by amplification - with primer-  
XX PT blocking oligo:nucleotide duplex(es) labelled with donor and  
XX PT acceptor fluorophore(s), to reduce non-specific priming

XX PS Example 4; Page 23; 41pp; English.

XX The presence of a blocking oligonucleotide partially complementary  
XX to an amplification primer in a PCR mixture reduces the number of  
XX non-specific priming events. When labelled with a fluorophore, the  
XX blocking oligonucleotide can also be used to monitor the amplification  
XX process by participating in fluorescence energy transfer. This  
XX energy transfer can be enhanced by using the blocking oligonucleotide  
XX as a template for ligation of its complementary sequence to the  
XX primer. In an example, the human acute myeloid leukaemia breakpoint  
XX related sequence was asymmetrically amplified using an excess primer  
XX and a limiting primer (see AAT09719 and AAT09720, respectively). The  
XX amplification process could be monitored using either a primer:  
XX blocking oligonucleotide duplex or a universal detection duplex  
XX coupled to a primer.

SQ Sequence 21 BP; 2 A; 6 C; 11 G; 2 T; 0 other;

Query Match

66.0%; Score 13.2; DB 17; Length 21;

Best Local Similarity 83.3%; Pred. No. 1.2e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gcgcgtgaagcgagc 19  
||| ||||| |||  
Db 1 gcgcgtgaagcgagc 18

## RESULT 6

AAT45484  
ID AAT45484 standard; DNA; 21 BP.

AC AAT45484;  
XX

XX 02-APR-1997 (first entry)  
XX

XX Human leukaemia breakpoint related sequence, limiting primer.  
XX

KW Detection duplex; asymmetric amplification; human; myeloid;  
KW leukaemia; breakpoint; related sequence; AMP-1; X chromosome;  
KW specific amelogenin; AMG-X; primer; excess; limiting; ss.  
XX

OS Synthetic.  
XX

XX US5567383-A.  
PN

XX 22-OCT-1996.  
PD

XX 16-DEC-1991; 91US-0808463.  
XX

XX 26-MAY-1994; 94US-0250849.  
PR

XX 16-DEC-1991; 91US-0808463.  
PR

XX (BIOT-) BIOTRONICS CORP.  
PA

XX Wang CJ, Wu K;  
PI

XX WPI; 1997-010704/01.  
DR

XX Nucleic acid detection by amplification - in presence of  
PT primer-complementary oligo:nucleotide to block non-specific priming  
XX

PS Example 4; Column 12; 17pp; English.  
XX

CC In an example to demonstrate the application of a detection duplex  
CC to monitor the asymmetric amplification of target nucleic acids,  
CC the human acute myeloid leukaemia breakpoint related sequence,  
CC (AMP-1), and the human X chromosome specific amelogenin (AMG-X)  
CC were used as amplification targets. The sequences of the primers  
CC used for AML-1 and AMG-X respectively were AAT45479 and AAT45480,  
CC duplex comprised AAT45481 and AAT45482 and the sequences of  
CC excess and limiting primers for AML-1 and AMG-X respectively were  
CC AAT45483 and AAT45484, and AAT45485 and AAT45486. Male human genomic DNA  
CC sequences were asymmetrically amplified with excess primer for  
CC 20 cycles prior to the addition of the primer duplex. The results  
CC compiled from detection at cycle 30 demonstrated the quantitative  
CC resolution and detection sensitivity by the duplex. The decrease  
CC in fluorescence intensity reflected the initial target doses.  
XX

SQ Sequence 21 BP; 2 A; 6 C; 11 G; 2 T; 0 other;  
XX

## Query Match

Best Local Similarity 66.0%; Score 13.2; DB 18; Length 21;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gcgcgtgaagcgagc 19  
||| ||||| |||  
Db 1 gcgcgtgaagcgagc 18

## RESULT 7

AAV09361

ID AAV09361 standard; DNA; 21 BP.

XX AAV09361;  
AC

XX 15-MAY-1998 (first entry)  
DT

XX Limiting primer used for detecting target nucleic acid AML-1.  
XX

XX Target; DNA duplex; donor; acceptor; fluorescent label; fluorophore;  
KW energy transfer; polymerase; primer; acute myeloid leukaemia; AML-1; ss.  
XX

OS Synthetic.  
OS

XX Homo sapiens.  
XX

XX US5712386-A.  
PN

XX 27-JAN-1998.  
PD

XX 04-MAY-1995; 95US-0434474.  
XX

XX 26-MAY-1994; 94US-0250849.  
PR

XX 16-DEC-1991; 91US-0808463.  
PR

XX 04-MAY-1995; 95US-0434474.  
PR

XX (BIOT-) BIOTRONICS CORP.  
PA

XX Wang CJ, Wu K;  
PI

XX WPI; 1998-120033/11.  
DR

XX Kits for detecting target nucleic acids - and DNA duplexes with  
PT donor and acceptor fluorescent labels  
XX

PS Example 4; Column 12; 17pp; English.  
XX

CC This is a limiting primer used in a kit for detecting a target nucleic  
CC acid of human acute myeloid leukaemia breakpoint related sequence  
CC (AML-1). The kit is a DNA duplex which comprises a first oligonucleotide  
CC capable of acting as a primer, with or without a segment noncontiguous to  
CC its priming sequence, for use with a polymerase in the amplification of a  
CC target nucleic acid, a second oligonucleotide which is hybridised, via  
CC at least 5 consecutive fully complementary nucleotide pairings, with the  
CC first oligonucleotide, the second oligonucleotide being incapable of  
CC acting as a primer for the polymerase, and a first fluorophore covalently  
CC attached to the first oligonucleotide, with one of the two fluorophores  
CC being a donor fluorophore and the other being an acceptor fluorophore, so  
CC that when the two fluorophores are in close proximity resonance energy  
CC transfer between them is allowed. Each of the first oligonucleotide and  
CC the second oligonucleotide contains 10--50 nucleotides. Another kit  
CC claimed comprises a first and second primer both optionally having a  
CC segment non-contiguous to a first or second priming sequence,  
CC respectively, which are used with a polymerase for the amplification of  
CC the target nucleic acid and an oligonucleotide which is incapable of  
CC acting as a primer for the polymerase and has at least 5 consecutive  
CC nucleotides fully complementary to at least 5 consecutive nucleotides of  
CC the first primer. Each of the first primer, the second primer and the  
CC oligonucleotide contains 10-50 nucleotides. A third kit for detecting a  
CC target nucleic acid comprises a first oligonucleotide being incapable of  
CC acting as a primer for use with a polymerase in the amplification of a  
CC target nucleic acid, and containing 10-50 nucleotides with a first  
CC fluorophore covalently attached to it, and a second oligonucleotide  
CC containing 5-30 nucleotides with a second fluorophore covalently attached  
CC to it, the second oligonucleotide having a free 3' OH and being capable  
CC of hybridizing, via at least 5 consecutive fully complementary nucleotide  
CC pairings, with the first oligonucleotide. The first oligonucleotide has  
CC an overhang beyond the 3' end of the second oligonucleotide by 1-12  
CC nucleotides, and the first and second fluorophores, one of which is a  
CC donor fluorophore and the other an acceptor fluorophore are in close  
CC proximity when the first oligonucleotide hybridises to the second  
CC oligonucleotide to allow resonance energy transfer between them. The  
CC kits are used in homogeneous assays in which the target nucleic acid  
CC sequence is amplified and the amplified target is detected without



CC conducting a separation step.

XX Sequence 21 BP; 2 A; 6 C; 11 G; 2 T; 0 other;

SQ

Query Match 66.0%; Score 13.2; DB 19; Length 21;

Best Local Similarity 83.3%; Pred. No. 1.2e+03;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gcgcgtgaagcgaagc 19

III IIIIIIIII III

Db 1 gcgcgtgaagcgcgcgc 18

RESULT 8

AAF95308

ID AAF95308 standard; DNA; 21 BP.

XX

AC AAF95308;

XX

DT 06-JUN-2001 (first entry)

XX

DE Human gene single nucleotide polymorphism #69.

XX

KW Human; variant thrombospondin 1; variant thrombospondin 4; SNP;

KW polymorphism; vascular disease; coronary artery disease; forensics;

KW myocardial infarction; atherosclerosis; stroke; venous thromboembolism;

KW pulmonary embolism; paternity test; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Variation replace(11,A)

FT /tag- a

FT /standard\_name= "single nucleotide polymorphism"

XX

PN WO200118250-A2.

XX

PD 15-MAR-2001.

XX

PF 07-SEP-2000; 2000WO-US24503.

XX

PR 10-SEP-1999; 99US-0153357.

XX

PR 26-JUL-2000; 2000US-0220947.

XX

PR 16-AUG-2000; 2000US-0225724.

XX

PA (WHEED ) WHITEHEAD INST BIOMEDICAL RES.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Lander ES, Gargill M, Ireland JS, Bolk S, Daley GQ, McCarthy JJ;

XX

DR WPI; 2001-226749/23.

XX

PT Nucleic acids comprising single nucleotide polymorphisms, useful in

PT applications such as forensics, paternity testing, medicine, genetic

PT analysis and phenotype correlations to diseases such as diabetes and

PT atherosclerosis

XX

PS Examples; Page 51; 242pp; English.

XX

CC The present invention provides a method of diagnosing a vascular disease

CC in an individual, involving determining the sequence at various

CC polymorphic sites within the human thrombospondin 1 and thrombospondin 4

CC genes. The sequences at a number of polymorphic sites are also provided

CC in the specification. In particular, the method can be used in the

CC diagnosis of atherosclerosis, myocardial infarction, coronary heart

CC disease, stroke, peripheral vascular diseases, venous thromboembolism

CC and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also

CC useful in forensics, paternity testing, genetic analysis and phenotype

CC correlations to diseases. The present sequence is an example of one of

CC the human gene SNPs shown in the specification.

XX

SQ Sequence 21 BP; 7 A; 5 C; 5 G; 4 T; 0 other;

Query Match 64.0%; Score 12.8; DB 22; Length 21;

Best Local Similarity 87.5%; Pred. No. 1.9e+03;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ccgtgaagcgaagcga 20

IIIIIIIIII I

Db 2 ccgtgaagcgaatgaa 17

RESULT 9

AAZ43694

ID AAZ43694 standard; DNA; 33 BP.

XX

AC AAZ43694;

XX

DT 23-FEB-2000 (first entry)

XX

DE Mass spectrometric mutation analysis primer 4.

XX

KW Primer; mass-spectrometry; genetic mutation; amplification; ss.

XX

OS Synthetic.

XX

PN DEL9824280-Al.

XX

PD 02-DEC-1999.

XX

PF 29-MAY-1998; 98DE-1024280.

XX

PR 29-MAY-1998; 98DE-1024280.

XX

PA (BRUK-) BRUKER DALTONIK GMBH.

XX

DR WPI; 2000-073581/07.

XX

PS Mass-spectrometric analysis of known gene mutations

XX

Example; Page 8; 16pp; German.

XX

CC This invention describes a method for mass-spectrometric analysis of

CC known genetic mutations, using modified nucleoside triphosphates to

CC improve the performance. The method comprises: (1) amplifying a DNA

CC sequence by polymerase chain reaction (PCR) using primers selected to

CC amplify a sequence containing the mutation; (2) adding a particular set

CC of modified nucleoside triphosphates (NTPs) to effect limited extension

CC of already present or newly added primers, where: (a) the extension

CC reaction stops at the next occurrence of a particular base in the DNA

CC strand being copied; (b) the extension reaction proceeds up to or past

CC the mutation site, so that wild-type amplification products will have a

CC different molecular weight from mutant amplification products; and

CC (c) the modification of the NTPs results in stabilization of the DNA

CC chains during ionization, a reduction in ion adduct formation, an

CC increase in ionization yields and/or a change in the mass of the DNA

CC chains; (3) performing the limited primer extension using an enzyme that

CC generates the complement of the DNA strand being copied; (4) performing

CC at least partial primer degradation and optionally further modification

CC of the amplification products; and (5) determining the mass of the

CC modified amplification products by mass spectrometry and assigning the

CC masses to wild type or mutant. AAZ43691-243717 represent primers used in

CC the method of the invention.

XX

SQ Sequence 33 BP; 6 A; 11 C; 11 G; 5 T; 0 other;

Query Match 63.0%;

Best Local Similarity 78.9%; Score 12.6; DB 21; Length 33;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcgcgtgaagcgaagc 19

IIII IIIIIIIII

Db 10 ggcgcgcgtgaagcgaagc 28

```

XX PN WO200056929-A2.
XX PD 28-SEP-2000.
XX PF 17-MAR-2000; 200WO-US07133.
XX PR 19-MAR-1999; 99US-0125251.
XX PA (CORR ) CORNELL RES FOUND INC.
XX PA (LOU ) UNIV LOUISIANA STATE.
XX PA (PURD ) PURDUE RES FOUND.
XX PI Barany F, Day JP, Hammer RP, Bergstrom DE;
XX PI WPI; 2000-638269/61.
XX DR
XX PT Coupled polymerase chain reaction-restriction endonuclease
XX PT digestion-ligase detection reaction to identify low abundance sequences
XX PT differing by single-base changes, insertion or deletion from high
XX PT abundance sequence in target sequences -
XX PS Example 7; Fig 6A; 103pp; English.
XX CC The specification describes the use of coupled polymerase chain reaction
XX CC (PCR), restriction endonuclease digestion (RED) and ligase detection
XX CC reaction (LDR) to identify one or more low abundance sequences differing
XX CC by one or more single base changes, inserts or deletions, from a high
XX CC abundance sequence, in several target nucleotide sequences. The method
XX CC involves 3 PCR reaction phases, a RED phase and an LDR phase. The method
XX CC is used to identify one or more low abundance sequences. The method is
XX CC also useful for detecting a wide variety of infectious diseases caused
XX CC by bacterial, viral, parasite and fungal infectious agent. Cancers can
XX CC also be detected by this method. The method is also used for detection,
XX CC identification and monitoring of pathogenic and indigenous municipal
XX CC waste water purification system and water reservoirs or in polluted
XX CC areas undergoing bioremediation and to detect plasmids containing genes
XX CC that can metabolise xenobiotics, to monitor specific target
XX CC microorganisms in population dynamic studies, or either to detect,
XX CC identify, or monitor specific target microorganisms modified
XX CC microorganisms in the environment and in industrial plants. The
XX CC present sequence represents a synthetic duplex DNA template, used in
XX CC the course of the invention.
XX SQ Sequence 50 BP; 11 A; 18 C; 12 G; 9 T; 0 other;

Query Match 63.0%; Score 12.6; DB 21; Length 50;
Best Local Similarity 78.9%; Pred. No. 2.4e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggccgcgtgaagcgaagc 19
   ||||| ||||| |||||
Db 12 ggccgcgtgaagcgaagc 30

RESULT 12
AAV95266/c
ID AAV95266 standard; RNA; 17 BP.
XX AC AAV95266;
XX DT
XX DT 24-FEB-1999 (first entry)
XX DE Human c-fos target sequence nucleotide position 97.
XX KW Human; c-fos; hammerhead ribozyme; hairpin ribozyme; target site;
XX KW cancer; oncogene; leukaemia; neuroblastoma; diagnosis; genetic drift;
XX KW mutation; diseased cell; ss.
XX OS Homo sapiens.
XX PN WO9832846-A2.

```

```

XX PN WO200056929-A2.
XX PD 28-SEP-2000.
XX PF 17-MAR-2000; 200WO-US07133.
XX PR 19-MAR-1999; 99US-0125251.
XX PA (CORR ) CORNELL RES FOUND INC.
XX PA (LOU ) UNIV LOUISIANA STATE.
XX PA (PURD ) PURDUE RES FOUND.
XX PI Barany F, Day JP, Hammer RP, Bergstrom DE;
XX PI WPI; 2000-638269/61.
XX DR
XX PT Coupled polymerase chain reaction-restriction endonuclease
XX PT digestion-ligase detection reaction to identify low abundance sequences
XX PT differing by single-base changes, insertion or deletion from high
XX PT abundance sequence in target sequences -
XX PS Example 7; Fig 6A; 103pp; English.
XX CC The specification describes the use of coupled polymerase chain reaction
XX CC (PCR), restriction endonuclease digestion (RED) and ligase detection
XX CC reaction (LDR) to identify one or more low abundance sequences differing
XX CC by one or more single base changes, inserts or deletions, from a high
XX CC abundance sequence, in several target nucleotide sequences. The method
XX CC involves 3 PCR reaction phases, a RED phase and an LDR phase. The method
XX CC is used to identify one or more low abundance sequences. The method is
XX CC also useful for detecting a wide variety of infectious diseases caused
XX CC by bacterial, viral, parasite and fungal infectious agent. Cancers can
XX CC also be detected by this method. The method is also used for detection,
XX CC identification and monitoring of pathogenic and indigenous municipal
XX CC waste water purification system and water reservoirs or in polluted
XX CC areas undergoing bioremediation and to detect plasmids containing genes
XX CC that can metabolise xenobiotics, to monitor specific target
XX CC microorganisms in population dynamic studies, or either to detect,
XX CC identify, or monitor specific target microorganisms modified
XX CC microorganisms in the environment and in industrial plants. The
XX CC present sequence represents a synthetic duplex DNA template, used in
XX CC the course of the invention.
XX SQ Sequence 50 BP; 11 A; 18 C; 12 G; 9 T; 0 other;

Query Match 63.0%; Score 12.6; DB 21; Length 50;
Best Local Similarity 78.9%; Pred. No. 2.4e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggccgcgtgaagcgaagc 19
   ||||| ||||| |||||
Db 12 ggccgcgtgaagcgaagc 30

RESULT 12
AAV95266/c
ID AAV95266 standard; RNA; 17 BP.
XX AC AAV95266;
XX DT
XX DT 24-FEB-1999 (first entry)
XX DE Human c-fos target sequence nucleotide position 97.
XX KW Human; c-fos; hammerhead ribozyme; hairpin ribozyme; target site;
XX KW cancer; oncogene; leukaemia; neuroblastoma; diagnosis; genetic drift;
XX KW mutation; diseased cell; ss.
XX OS Homo sapiens.
XX PN WO9832846-A2.

```

XX 30-JUL-1998.  
 XX 20-JAN-1998; 98WO-US01017.  
 PF 23-JAN-1997; 97US-0037658.  
 PR (RIBO-) RIBOZYME PHARM INC.  
 XX Jarvis T, McSwiggen JA, Stinchcomb DF;  
 XX WPI; 1998-427942/36.  
 XX Enzymatic nucleic acid molecules which specifically cleave RNA  
 PT derived from a c-fos gene - useful for treating conditions related  
 PT to levels of c-fos, especially cancer  
 XX  
 PS Claim 2; Page 50; 72pp; English.  
 XX The present invention describes an enzymatic nucleic acid molecule which  
 CC specifically cleaves RNA derived from a c-fos gene. AAV95401 to AAV95540  
 CC and AAV95541 to AAV95584 represent hammerhead ribozymes and hairpin  
 CC ribozymes, respectively, which specifically cleave human c-fos. AAV95261  
 CC to AAV95400 and AAV95385 to AAV95628 represent human c-fos target  
 CC sequences. The enzymatic nucleic acid molecules can be used for treating  
 CC cancer associated with elevated levels of c-fos oncogene, especially  
 CC leukaemias, neuroblastomas and lung, breast and colon cancers. The  
 CC ribozymes may also be used as diagnostic tools to examine genetic drift  
 CC and mutations within diseased cells, or, to detect the presence of c-fos  
 CC RNA in a cell.  
 XX  
 SQ Sequence 17 BP; 2 A; 8 C; 3 G; 4 U; 0 other;  
 XX

Query Match 61.0%; Score 12.2; DB 19; Length 17;  
 Best Local Similarity 82.4%; Pred. No. 3.6e+03;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 ggcgcctgaagcgaag 17  
 Db 17 GGCCTGTGAAGCAGAG 1  
 ||||| ||||| ||  
 ||||| ||||| ||

RESULT 13  
 AAA84390  
 ID AAA84390 standard; DNA; 19 BP.  
 XX  
 AC AAA84390;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE Cyclin D3 ribozyme binding site #1.  
 XX  
 KW Ribozyme; hairpin; hammerhead; gene therapy; vasotropic;  
 KW restenosis; ss.  
 XX  
 OS Mammalia.  
 XX  
 XX  
 XX W0200032765-A2.  
 PN  
 PD 08-JUN-2000.  
 XX  
 XX  
 XX 06-DEC-1999; 99WO-US28772.  
 PF  
 XX  
 PR 04-DEC-1998; 98US-0110954.  
 XX  
 XX (IMMU-) IMMUSOL INC.  
 PA  
 XX Tritz R, Welch PJ, Barber JR, Robbins JM;  
 XX WPI; 2000-412314/35.  
 XX  
 XX New hairpin and hammerhead ribozyme for inhibiting restenosis, cleaves

PT RNA encoding a cyclin or cell-cycle dependent kinase other than CDK1,  
 PT PCNA and Cyclin B1 -  
 XX  
 PS Disclosure; Page 76; 109pp; English.  
 XX  
 CC The present invention relates to a hairpin or hammerhead ribozyme,  
 CC designed to cleave RNA encoding a cyclin or cell-cycle dependent kinase  
 CC other than cell-cycle dependent kinases CDK1, PCNA and Cyclin B1.  
 CC Representative examples of ribozyme recognition sites are given in  
 CC AAA82415 to AAA86787. The ribozyme of the invention is useful for  
 CC inhibiting restenosis by introduction of the ribozyme into cells.  
 CC The ribozyme is resistant to endonuclease activity and hence is  
 CC efficient in restenosis treatment.  
 XX  
 SQ Sequence 19 BP; 3 A; 4 C; 7 G; 5 T; 0 other;  
 XX

Query Match 61.0%; Score 12.2; DB 21; Length 19;  
 Best Local Similarity 82.4%; Pred. No. 3.6e+03;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 gccgtgaagcgaagca 20  
 Db 3 gctgtgttcgaagca 19  
 ||| ||| ||||| |||  
 ||| ||| ||||| |||

RESULT 14  
 AAF74126/C  
 ID AAF74126 standard; DNA; 20 BP.  
 XX  
 AC AAF74126;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Primer #60.  
 XX  
 KW Solute carrier family 6 neurotransmitter transporter; seotonin 4;  
 KW SLC6A4; genotyping; allele specific oligonucleotide; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W0200109161-A1.  
 PN  
 PD 08-FEB-2001.  
 XX  
 XX 31-JUL-2000; 2000WO-US20638.  
 PF  
 XX  
 XX 29-JUL-1999; 99US-0146290.  
 PR  
 XX (GENA-) GENAISSANCE PHARM INC.  
 PA  
 XX Denton RR, Duda A, Nandabalan K, Sanchis A, Stephens JC;  
 PI  
 XX WPI; 2001-123317/13.  
 DR  
 XX  
 XX New isolated polynucleotide comprising a polymorphic variant for the  
 PT solute carrier family 6 neurotransmitter transporter, serotonin member  
 PT 4 gene for identifying drugs for treating disorders related to  
 PT expression of the protein -  
 XX  
 XX Example 1; Page 37; 152pp; English.  
 PS  
 XX  
 XX The present invention relates to a polymorphic variant of a reference  
 CC sequence for the solute carrier family 6 neurotransmitter  
 CC transporter, serotonin member 4 (SLC6A4) gene or a fragment of it  
 CC or a sequence complementary to the first sequence.  
 CC The invention is used in producing a recombinant organism  
 CC that can be used to express SLC6A4 for protein structure analysis and  
 CC binding studies. A composition comprising a genotyping oligonucleotide  
 CC is used to detect a polymorphism in the SLC6A4 gene.  
 XX  
 SQ Sequence 20 BP; 2 A; 8 C; 3 G; 7 T; 0 other;  
 XX

Query Match 61.0%; Score 12.2; DB 22; Length 20;  
Best Local Similarity 82.4%; Pred. No. 3.6e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 gccgtgaagcggaaggca 20  
||| ||||| |||||  
Db 19 GCGGTGAAATGAAGGCA 3

## RESULT 15

AAF69762  
ID AAF69762 standard; DNA; 22 BP.

XX  
AC AAF69762;

XX  
DT 18-APR-2001 (first entry)

XX  
DE Human IL4Ralpha gene PCR primer #98.

XX  
KW Polymorphism; human; interleukin 4 receptor-alpha; IL4R-alpha;  
allergic disease; PCR primer; ss.

XX  
OS Homo sapiens.

XX  
PN WO200104270-A1.

XX  
PD 18-JAN-2001.

PF 13-JUL-2000; 2000WO-US19094.

XX  
PR 13-JUL-1999; 99US-0143435.

XX  
PA (GENA-) GENAISSANCE PHARM INC.

XX  
PI Chew A, Denton RR, Duda A, Nandabalan K, Stephens JC;

XX  
PI Windemuth AK;

XX  
DR WPI; 2001-103078/11.

XX  
PT New isolated polynucleotide useful for the identification of  
therapeutics in allergic diseases is new -

XX  
PS Example 1; Page 65; 188pp; English.

XX  
CC The present invention relates to polymorphisms of the human interleukin 4  
receptor-alpha gene (IL4R-alpha; see AAF57718 for the reference  
sequence). Polynucleotides comprising polymorphic gene variants are  
useful for therapeutic purposes. For example, where a patient may benefit  
from expression of a particular IL4Ralpha protein isoform, an expression  
vector encoding the isoform may be administered to the patient. It may  
be desirable to decrease or block expression of a particular IL4Ralpha  
isoform, which may be done by turning off by transforming a targeted  
organ, tissue or cell population with an expression vector that expresses  
high levels of untranslatable mRNA for the isoform. Specific therapeutics  
identified by these methods may be useful for allergic diseases. The  
present sequence is a PCR primer for human IL4R-alpha.

XX  
SQ Sequence 22 BP; 9 A; 2 C; 10 G; 1 T; 0 other;

Query Match 61.0%; Score 12.2; DB 22; Length 22;  
Best Local Similarity 82.4%; Pred. No. 3.7e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 gccgtgaagcggaaggca 20  
||| ||||| |||||  
Db 5 gcagggaagagaaggca 21

Search completed: October 2, 2001, 16:18:35  
Job time: 15479 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 12:00:36 ; Search time 876.95 Seconds  
(without alignments)  
14.320 Million cell updates/sec

Title: US-09-757-100B-3

Perfect score: 20

Sequence: 1 ccgcgggtcacagtgtcg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SIDSB/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSB/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSB/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	22	AAC65535 Human focal adhesi
2	15	75.0	15	22	AAC65535 Human focal adhesi
3	13.8	69.0	31	22	AAC83406 Oligonucleotide sf
4	13.6	68.0	41	21	AZ45393 PCR primer used to
5	13.2	66.0	29	21	AAA04346 Polymorphic fragme
6	13.2	66.0	31	19	AAV42721 3' PCR primer used
7	13.2	66.0	39	21	AAA37483 Arabidopsis thalia
8	13.2	66.0	50	21	AAC61924 PCR primer for DNA
9	12.8	64.0	39	11	AAQ04938 Oligonucleotide ca
10	12.8	64.0	44	21	AAA72106 HIV-1 gp41 C-helic
11	12.6	63.0	29	19	AAV59089 Primer for ribonuc

c	12	12.6	63.0	32	21	AAZ35374	Sense primer for
13	12.6	63.0	47	20	AAZ01125	Probe for human PG	
14	12.2	61.0	18	21	AAA38381	Human Ets-2 phosph	
c	15	12.2	61.0	29	21	AAZ29338	Primer V-beta-3 fo
c	16	12.2	61.0	30	20	AAZ10873	PCR primer for the
17	12.2	61.0	31	18	AAZ63364	Primer murf2. Syn	
18	12.2	61.0	31	20	AAX86099	PCR primer used to	
19	12.2	61.0	39	18	AAT63369	Primer murf4. Syn	
20	12.2	61.0	39	20	AAX86104	PCR primer used to	
21	12.2	61.0	40	21	AAX55715	Anabaena variabili	
22	12.2	61.0	41	21	AAC61918	PCR primer for DNA	
23	12.2	61.0	42	21	AAA37487	Arabidopsis thalia	
24	12.2	61.0	43	21	AAA37503	Yeast acyltransfer	
25	12.2	61.0	44	21	AAA37489	Arabidopsis thalia	
26	12.2	61.0	44	21	AAA37507	Yeast acyltransfer	
27	12.2	61.0	45	21	AAC61920	PCR primer for DNA	
28	12.2	61.0	50	21	AAZ47143	Liposome membrane-	
c	29	12	60.0	41	19	AAV35598	Chicken IgG light
30	12	60.0	42	21	AAA37485	Arabidopsis thalia	
31	12	60.0	43	18	AAT66939	Cels dockerin gene	
32	12	60.0	46	21	AAA37481	Arabidopsis thalia	
c	33	11.8	59.0	16	18	AAT93016	M. tuberculosis 16
c	34	11.8	59.0	20	19	AAV45725	Human NKCC2 gene e
35	11.8	59.0	24	20	AAX36171	PCR primer used to	
36	11.8	59.0	29	18	AAT92996	M. tuberculosis 16	
37	11.8	59.0	30	18	AAT92997	M. tuberculosis 16	
38	11.8	59.0	30	18	AAT92998	M. tuberculosis 16	
39	11.8	59.0	35	17	AAT10707	Mycobacterium genu	
40	11.8	59.0	36	14	AAC38115	Mycobacterium 16S	
c	41	11.8	59.0	40	20	AAX99220	M. tuberculosis 16
c	42	11.8	59.0	40	20	AAX99221	M. bovis 16S rRNA
c	43	11.8	59.0	40	20	AAX99225	M. smegmatis 16S r
c	44	11.8	59.0	40	20	AAX99227	M. intracellulare
c	45	11.8	59.0	44	17	AAT06805	Mycobacteria 16S r

#### ALIGNMENTS

RESULT 1

AAC65535  
ID AAC65535 standard; DNA; 20 BP.

XX AC AAC65535;

XX AC AAC65535;

XX DT 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #1.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer;

XX embryonic development disorder; angiogenic disorder; wound healing;

XX antisense; phosphorothioate; ss.

XX Homo sapiens.

XX US6133031-A.

XX PD 17-OCT-2000.

XX PF 19-AUG-1999; 99US-0377310.

XX PR 19-AUG-1999; 99US-0377310.

XX (ISIS-) ISIS PHARM INC.

XX Monia BP, Gaarde WA;

XX WPI; 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase

XX expression, especially useful for inhibiting retinal

XX neovascularization, or for diagnosing and treating e.g. colon cancer

XX







KW human; furin; ablation therapy; ss.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX WO9830709-A2.  
 PN 16-JUL-1998.  
 PD 14-JAN-1998; 98WO-US00715.  
 PF 13-JAN-1998; 98US-0038339.  
 XX 14-JAN-1997; 97US-0035473.  
 PR 27-FEB-1997; 97US-0038339.  
 XX (CHAD//) CHADA S.  
 PA (JOLLY//) JOLLY D J.  
 PA (MOOR//) MOORE M D.  
 XX Chada S, Jolly DJ, Moore MD;  
 PI WPI; 1998-399153/34.  
 XX Non-immunogenic pro-drug activating enzyme(s) and selectable  
 PT marker(s) - are used in gene therapy for the treatment of a wide  
 PT variety of disorder(s)  
 XX Example 8; Page 53; 121pp; English.  
 PS PCR primers AAV42720-21 were used to amplify human furin cDNA (see  
 CC AAV42731). The amplified product is used in the retroviral vector of the  
 CC invention to encode a cell bound prodrug convertase for ablation therapy.  
 CC The specification describes a gene delivery vehicle which directs  
 CC expression of a non-immunogenic selectable marker or molecule which is  
 CC capable of activating a previously inactive compound. Vectors expressing  
 CC the markers and a heterologous sequence are useful in gene therapy. The  
 CC vectors can be used to deliver a molecule into a target area where it may  
 CC cause the activation of a previously inactive substance.  
 XX Sequence 31 BP; 5 A; 12 C; 10 G; 4 T; 0 other;  
 SQ

Query Match 66.0%; Score 13.2; DB 19; Length 31;  
 Best Local Similarity 83.3%; Pred. No. 1.le+03;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 gcggggtcacagtgtcg 20  
 Db 13 gtggggtcacagagggcg 30

RESULT 7  
 AAA37483  
 ID AAA37483 standard; DNA; 39 BP.  
 AC AAA37483;  
 XX 15-AUG-2000 (first entry)  
 DT Arabidopsis thaliana acyltransferase ATAT7 PCR primer, SEQ ID NO:173.  
 DE Arabidopsis thaliana.  
 XX Acyltransferase; lipid synthesis; recombinant expression;  
 KW membrane fluidity; cold resistance; transgenic plant;  
 KW baculovirus expression; PCR primer; ss.  
 XX Arabidopsis thaliana.  
 OS WO200018889-A2.  
 PN 06-APR-2000.  
 PD 24-SEP-1999; 99WO-US22231.  
 PF Savidge B, Lassner MW, Weiss JD, Post-Beittenmiller D;  
 XX (CALJ ) CALGENE LLC.

PR 25-SEP-1998; 98US-0101939.  
 XX (CALJ ) CALGENE LLC.  
 XX Lassner MW, Emig RA, Ruezinsky DM, Van Eenennaam A;  
 PI WPI; 2000-303447/26.  
 DR Novel acyltransferase related proteins useful for altering membrane  
 PT fluidity in plant cells e.g. to induce chill tolerance -  
 XX Example 7; Page 31; 126pp; English.  
 PS The invention relates to nucleic acids encoding novel plant  
 CC acyltransferase-like proteins (AAA37343-A37445) which comprise one of 8  
 CC conserved acyltransferase motifs (AA199474-Y99481). Acyltransferases  
 CC catalyse the transfer of acyl groups from a donor to a variety of  
 CC substrates such as glycerides, sterols, stanols and phosphatides.  
 CC Such enzymes play a key role in lipid synthesis, and thereby affect the  
 CC characteristics of the plant. For example, cold-hardened plants have  
 CC different lipid concentrations in the cell membrane compared to  
 CC non-hardened plants, which makes the membrane more fluid and the plant  
 CC more tolerant of low temperatures. The nucleic acid sequences of the  
 CC invention can be used as probes or for expressing acyltransferase-like  
 CC proteins in host cells e.g., for recombinant protein production. They  
 CC may be expressed in plant cells to alter the lipid composition of the  
 CC plant e.g., for the production of chill-resistant plants, or for altering  
 CC the composition of plant oils. Sequences AAA37473-A37494 represent PCR  
 CC primers used in an exemplification of the invention to amplify the  
 CC Arabidopsis thaliana acyltransferase cDNAs (AAA37331-A37342) for cloning  
 CC into plant and baculovirus expression constructs.  
 XX Sequence 39 BP; 11 A; 9 C; 12 G; 7 T; 0 other;  
 SQ

Query Match 66.0%; Score 13.2; DB 21; Length 39;  
 Best Local Similarity 83.3%; Pred. No. 1.le+03;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ccgcggttcacagtgggt 18  
 Db 5 ccgcggttcacagtgggt 22

RESULT 8  
 AAC61924  
 ID AAC61924 standard; DNA; 50 BP.  
 XX AAC61924;  
 AC AAC61924;  
 XX 06-MAR-2001 (first entry)  
 DT PCR primer for DNA encoding a prenyltransferase designated ATP78.  
 DE Prenyltransferase; ATP71; ATP72; ATP73; ATP74; ATP75; ATP76; ATP77;  
 KW ATP78; ATP79; ATP10; ATP11; ATP12; tocopherol; homogenetic acid;  
 KW phytylpyrophosphate; 2-methyl-6phytylbenzoquinol; antioxidant;  
 KW nutritional supplement; PCR primer; ss.  
 XX Arabidopsis sp.  
 OS WO200063391-A2.  
 PN 26-OCT-2000.  
 PD 14-APR-2000; 2000WO-US10368.  
 PF 15-APR-1999; 99US-0129899.  
 XX 30-JUL-1999; 99US-0146461.  
 PR (CALJ ) CALGENE LLC.  
 PA Savidge B, Lassner MW, Weiss JD, Post-Beittenmiller D;  
 PI

XX WPI; 2000-647519/62.

XX An isolated nucleic acid sequence encoding prenyltransferase used to

PT transform plant cells to increase the production of tocopherols -

XX

XX Example 2; Page 26; 114pp; English.

XX PCR primers AAC61924-25 were used to amplify DNA encoding a

CC prenyltransferase. The specification describes prenyltransferases

CC designated ATP11, ATP2, ATP3, ATP4, ATP5, ATP6, ATP7, ATP8,

CC ATP9, ATP10, ATP11, and ATP12. The biosynthesis of alpha-tocopherol

CC in higher plants involves the condensation of homogentisic acid and

CC phytylpyrophosphate to form 2-methyl-6phytylbenzoquinol, which can form

CC various tocopherols. The prenyltransferase polynucleotides are useful in

CC transforming host cells to alter the expression of prenyltransferase in

CC these cells. The transformed cells are used in the production of

CC tocopherols which are of use in the pharmaceutical industry as

CC antioxidants and also in the food industry as nutritional supplements.

XX

SQ Sequence 50 BP; 9 A; 16 C; 14 G; 11 T; 0 other;

Query Match 66.0%; Score 13.2; DB 21; Length 50;

Best Local Similarity 83.3%; Pred. No. 1.1e+03;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccgcgggctcacagtgt 18

||||| | |||| |

Db 5 ccgcggcgcacaatggt 22

RESULT 9

AAQ04938

ID AAQ04938 standard; DNA; 39 BP.

XX

AC AAQ04938;

XX

XX 24-OCT-1990 (first entry)

XX

DE Oligonucleotide carrying mutation for factor VIII gene.

XX

KW Human factor VIII analogue; ss.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT mutation 18..23

FT /\*tag= a

XX

XX WO9005530-A.

XX

PD 31-MAY-1990.

XX

XX 14-NOV-1989; 89WO-0005049.

XX

XX 14-NOV-1988; 88US-0270882.

XX

PA (GENE-) GENETICS INST INC.

XX

PI Kaufman RJ, Pittman DD;

XX

XX WPI; 1990-193265/25.

XX

XX New hybrid DNA encoding hybrid procoagulant proteins -

PT prep'd. by modifying DNA encoding human factor VIII.

XX

PS Disclosure; ; pp; English.

XX

XX Factor VIII analogue is sufficiently mutated from the original gene

CC that it is not recognised by blood Abs of the patient. The analogue

CC is composed of human FVIII but carries the B-domain of human FV in

CC place of the FVIII B-domain.

CC Oligonucleotides can be used to alter the profile of the blood

CC factor without significantly altering its activity.

XX

SQ Sequence 39 BP; 9 A; 11 C; 11 G; 8 T; 0 other;

Query Match 64.0%; Score 12.8; DB 11; Length 39;

Best Local Similarity 87.5%; Pred. No. 1.8e+03;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ccgcgggctcacagtgg 17

|||| | |||| |

Db 19 ccgcgtgtcacagtgg 34

RESULT 10

AAAY72106

ID AAA72106 standard; DNA; 44 BP.

XX

AC AAA72106;

XX

XX 24-NOV-2000 (first entry)

XX

DE HIV-1 gp41 C-helical domain 5' PCR primer.

XX

KW HIV-1; gp41; N-helical domain; heptad repeat region; C-helical domain;

KW gp41 transmembrane-proximal amphipathic alpha-helical segment;

KW core 6-helix bundle; viral entry inhibition; immunogenic;

KW antibody; humoral response; broad spectrum vaccine; anti-HIV;

KW envelope glycoprotein; prophylaxis; therapy; PCR primer; ss.

XX

OS Human immunodeficiency virus type 1.

XX

XX WO2000040616-A1.

XX

PD 13-JUL-2000.

XX

XX 10-JAN-2000; 2000WO-US000456.

XX

XX 08-JAN-1999; 99US-0115404.

PR 07-JAN-2000; 2000US-0480336.

XX

PA (WILD/) WILD C T.

PA (WEIS/) WEISS C D.

XX

PI Wild CT, Weiss CD;

XX

DR WPI; 2000-465959/40.

XX

PT Raising neutralizing antibody response to human immunodeficiency virus,

PT comprises administering a polypeptide capable of forming a stable

PT coiled-coil solution structure -

XX

PS Example 5; Page 54; 97pp; English.

XX

CC The invention relates to raising a neutralising antibody response to a

CC broad spectrum of HIV (human immunodeficiency virus) strains and

CC isolates, comprising the administration of a peptide which corresponds

CC to or mimics highly conserved portions of the gp41 envelope glycoprotein

CC which are important in mediating the process of viral entry into host

CC cells. Such peptides can correspond to or mimic the coiled coil

CC solution structure of the N-helical domain (the heptad repeat

CC region), or can correspond or mimic the C-helical domain (the

CC transmembrane-proximal amphipathic alpha-helical segment), or the

CC gp41 core 6-helix bundle, which is formed by the interaction of

CC the N- and C-helical domains of three gp41 proteins. The peptides

CC can be administered either singly or as a combination (particularly

CC a combination of N-helical and C-helical peptides), and can be

CC multimerised. For example, N- and C-helical domain peptides can be

CC alternately linked together to form a peptide which mimics the core

CC 6-helix bundle. Administration of the peptide(s) generates a humoral

CC response, with the production of antibodies against gp41 structures

CC involved in viral entry. As these portions of gp41 are well conserved,

CC such antibodies may be effective against a broad range of HIV strains  
 CC and isolates. The peptide compositions may be administered as a  
 CC prophylactic or therapeutic vaccine to generate antibodies which reduce  
 CC or inhibit the ability of HIV to infect uninfected cells. A composition  
 CC comprising polyclonal or monoclonal antibodies can be administered to  
 CC reduce HIV infection of uninfected cells. Antibodies raised against  
 CC entry-relevant gp41 structures may also be used therapeutically and as  
 CC tools to further elucidate the mechanism of HIV cell entry.  
 CC Sequences AAA72104-A72107 represent PCR primers used in an  
 CC exemplification of the invention to construct DNA (AAA72102) encoding a  
 CC protein construct (RAB14603) comprising the N- and C-helical domains of  
 CC HIV-1 gp41 separated by a short peptide linker.  
 XX  
 SQ Sequence 44 BP; 13 A; 10 C; 15 G; 6 T; 0 other;

Query Match 64.0%; Score 12.8; DB 21; Length 44;  
 Best Local Similarity 87.5%; Pred. No. 1.8e+03;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 2 cgcgggctcacagtgg 17  
 ||| ||||| |||||  
 Db 8 cgcgggctcacagtgg 23

RESULT 11  
 AAV59089  
 ID AAV59089 standard; DNA; 29 BP.  
 XX  
 AC AAV59089;  
 DT 15-JAN-1999 (first entry)  
 XX  
 DE Primer for ribonucleotide reductase R1 coding sequence.  
 XX  
 KW Ribonucleotide reductase R1; tumourigenicity modulation; therapy;  
 KW neoplastic cell; tumour growth; tumour regression; PCR primer; mouse; ss.  
 XX  
 OS Synthetic.  
 OS Mus sp.  
 XX  
 PN WO9841231-A1.  
 PD 24-SEP-1998.  
 XX  
 PF 18-MAR-1998; 98WO-CA00242.  
 XX  
 PR 02-OCT-1997; 97US-0060669.  
 PR 19-MAR-1997; 97US-0040837.  
 XX  
 PA (GENE-) GENESENSE TECHNOLOGIES INC.  
 XX  
 PI Wright JA, Young AH;  
 XX  
 DR WPI; 1998-520956/44.  
 XX  
 PT Use of ribonucleotide reductase R - for modulating the  
 PT tumourigenicity and metastatic potential of neoplastic cells in a  
 PT mammal, particularly for inhibiting tumour growth  
 XX  
 PS Example; Page 21; 56pp; English.  
 XX  
 CC This sequence represents a PCR primer for the mouse ribonucleotide  
 CC reductase R1 (RRR1) gene. The gene and the protein it encodes can be  
 CC used in the methods of the invention for modulating tumourigenicity of  
 CC neoplastic cells in a mammal. The RRR1 can slow tumour growth and cause  
 CC tumour regression and reduction of tumourigenicity and metastatic  
 CC potential.  
 XX  
 SQ Sequence 29 BP; 3 A; 11 C; 7 G; 8 T; 0 other;

Query Match 63.0%; Score 12.6; DB 19; Length 29;

Best Local Similarity 78.9%; Pred. No. 2.2e+03;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtggtc 19  
 ||| ||| ||||| |||||  
 Db 2 cctcgcgtgacagtcgctc 20

RESULT 12  
 AAZ35374/C  
 ID AAZ35374 standard; DNA; 32 BP.  
 XX  
 AC AAZ35374;  
 DT 27-MAR-2000 (first entry)  
 XX

DE Sense primer for methylthioadenosine phosphorylase gene exons 5-8.  
 XX  
 KW Chromosome 9p21; human; cancer; tumour; diagnosis; prognosis;  
 KW methylthioadenosine phosphorylase; MTPAP; glioma; melanoma;  
 KW primary lymphoid malignancy; non-small cell lung cancer;  
 KW head and neck cancer; ovarian cancer; bladder cancer;  
 KW chondrosarcoma; PCR primer; ss.  
 XX

OS Homo sapiens.  
 XX  
 PN WO9967634-A1.  
 XX  
 PD 29-DEC-1999.  
 XX

PF 21-JUN-1999; 99WO-US13981.  
 XX  
 PR 23-JUN-1998; 98US-0090411.  
 PR 17-JUN-1999; 99US-0335231.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX

PI Carson DA, Schmid M, Carrera CJ;  
 XX  
 DR WPI; 2000-126650/11.  
 XX

PT Diagnosing and determining prognosis for cancer causatively associated  
 PT with derangements of chromosome 9p21 -  
 XX  
 PS Claim 7; Page 16; 55pp; English.  
 XX

CC This sense primer is designed for the PCR amplification of exons  
 CC 5-8 of the human methylthioadenosine phosphorylase (MTPAP) gene on  
 CC chromosome 9p21 (see AAZ35374). The primer is used, with an  
 CC antisense primer (see AAZ35375), in a PCR amplification (505 bp  
 CC product) in step (a) of claimed methods for diagnosis of, and  
 CC determining a prognosis for, cancer causatively associated with  
 CC derangements of chromosome 9p21. Step (a) of these methods  
 CC involves determining whether any portion of the 9p21 chromosome  
 CC including and telomeric to STS 3.21 is deleted, or whether any  
 CC portion of the gene encoding MTPAP is deleted. Step (b) involves  
 CC determining whether any portion of 9p21 centromeric to STS 3.21  
 CC is deleted, or whether any portion of the gene coding for p16 is  
 CC deleted. The results are indicative of a cancer at an early stage  
 CC or advanced stage of tumour development. Primer pairs (see  
 CC AAZ35374-75) are provided for use in the claimed methods, where the  
 CC cancer is a glioma, primary lymphoid malignancy, non-small cell  
 CC lung cancer, melanoma, head and neck cancer, ovarian cancer,  
 CC bladder cancer or a chondrosarcoma (claimed).  
 XX

SQ Sequence 32 BP; 8 A; 9 C; 6 G; 9 T; 0 other;

Query Match 63.0%; Score 12.6; DB 21; Length 32;  
 Best Local Similarity 78.9%; Pred. No. 2.2e+03;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ccgcgggctcacagtggtc 19

Db 19 CTGAGGTCTCATAGTGTC 1  
 | | | | | | | | | | | | | | | |

## RESULT 13

AAZ01125  
 ID AAZ01125 standard; DNA; 47 BP.  
 XX  
 AC AAZ01125;  
 XX  
 DT 27-SEP-1999 (first entry)  
 DE  
 DE Probe for human PGI biallelic marker 4-50-293.  
 XX  
 KW PGI gene; biallelic marker; PCR primer; PGI-related biallelic marker;  
 KW cancer; prostate cancer; diagnosis; therapy; prostate specific antigen;  
 KW PSA; human; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO9932644-A2.  
 XX  
 XX 01-JUL-1999.  
 PD  
 XX 22-DEC-1998; 98WO-1B02133.  
 PF  
 XX 09-SEP-1998; 98US-0099658.  
 PR  
 XX 22-DEC-1997; 97US-0596306.  
 XX  
 XX (GEST ) GENSET.

PA Blumenfeld M, Bougueleret L, Chumakov I, Cohen D;  
 PI WPI; 1999-405178/34.  
 XX

PT Use of a prostate cancer associated gene and biallelic markers  
 PT derived from it

XX Claim 4; Page 328; 385pp; English.

CC The invention relates to a mammalian PGI gene and protein, and a set of  
 CC PGI biallelic markers. The PGI polynucleotide and biallelic markers are  
 CC used in a hybridisation assay, a sequencing assay, or in an  
 CC allele-specific amplification assay for determining the identity of a  
 CC nucleotide at a PGI-related biallelic marker. The methods can be used to  
 CC detect and to assess the risk of developing cancer or prostate cancer.  
 CC Early-stage diagnosis of prostate cancer relies on prostate specific  
 CC antigen (PSA) dosage. However, the effectiveness of this is limited due  
 CC to its inability to discriminate between malignant and non-malignant  
 CC affections of the organ. A need exists for both a reliable diagnostic  
 CC procedure which would enable early-stage diagnosis, and for preventative  
 CC and curative treatments of the disease. The PGI gene can be used for  
 CC detection of prostate cancer, and the risk of developing it in the  
 CC future, and can also be used to determine therapies for the disease.

XX Sequence 47 BP; 14 A; 9 C; 14 G; 10 T; 0 other;

Query Match 53.0%; Score 12.6; DB 20; Length 47;  
 Best Local Similarity 78.9%; Pred. No. 2.2e+03;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccgcgggctcacagtgtgc 19  
 | | | | | | | | | | | | | | | |  
 Db 10 cagggggctgacagtggac 28

## RESULT 14

AAA38381  
 ID AAA38381 standard; DNA; 18 BP.  
 XX  
 AC AAA38381;

XX  
 DT  
 DE  
 DE Human Ets-2 phosphorothioate antisense oligonucleotide, SEQ ID NO:40.  
 XX  
 KW Ets-2; human; transcription factor; chromosome 21q22.3; cancer; invasion;  
 KW metastasis; skeletal abnormality; Down's syndrome; expression inhibition;  
 KW phosphorothioate; antisense; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6054316-A.  
 XX  
 PD 25-APR-2000.  
 XX  
 PF 25-JUN-1999; 99US-0344579.  
 XX  
 PR 25-JUN-1999; 99US-0344579.  
 XX  
 XX (ISIS-) ISIS PHARM INC.  
 XX  
 PA Baker BF, Cowser LM;  
 PI WPI; 2000-338495/29.  
 XX

XX Antisense compound, 8-30 nucleobases in length, inhibiting the  
 PT expression Ets-2 is useful for treating cancer and detecting Ets-2  
 PT expression  
 XX  
 PS Claim 3; Column 40; 31pp; English.

XX Sequences AAA38349-A38388 represent antisense oligonucleotides targeted  
 CC to the human Ets-2 gene, which inhibit its expression. The antisense  
 CC oligonucleotides were designed to target different regions of the human  
 CC Ets-2 RNA, and were analysed for their effect on Ets-2 mRNA levels by  
 CC quantitative real-time PCR. The Ets-domain transcription factors are a  
 CC family of proteins which are involved in controlling key cellular events  
 CC such as proliferation, differentiation and development. The Ets domain  
 CC is a DNA-binding domain shared by all members of this family. Through  
 CC this motif, Ets family members bind to the promoter regions of various  
 CC genes at a GCA consensus sequence, thereby acting as either repressors  
 CC or activators of the gene. All but one Ets family protein bind to DNA as  
 CC a monomer. Ets-2 has been implicated in the regulation of cellular  
 CC proliferation and differentiation. The Ets-2 gene is located at  
 CC chromosome 21q22.3, which is within a region known to undergo  
 CC translocations associated with malignancies. Ets-2 has been found to be  
 CC upregulated in several cancers, including lymphoblastic leukaemia. It  
 CC may also play a role in the cancer phenotype, as it activates the  
 CC urokinase plasminogen activator (uPA) promoter and the promoters of  
 CC metalloproteinases in response to epidermal growth factor (EGF)  
 CC stimulation. High levels of uPA and metalloproteinases are associated  
 CC with tumour invasion and metastasis in breast cancers. As the Ets-2 gene  
 CC is located on chromosome 21, which is triplicated in Down's syndrome, it  
 CC is also thought to be responsible for the skeletal abnormalities present  
 CC in this condition. The antisense oligonucleotides of the invention are  
 CC useful for the treatment or prophylaxis of conditions associated with  
 CC Ets-2 expression, especially cancer.

XX Sequence 18 BP; 1 A; 5 C; 9 G; 3 T; 0 other;

Query Match 61.0%; Score 12.2; DB 21; Length 18;  
 Best Local Similarity 82.4%; Pred. No. 3.4e+03;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 gcggggtcacagtgtgc 19  
 | | | | | | | | | | | | | | | |  
 Db 2 gcggggtcacagtgtgc 18

## RESULT 15

AAA29338/c  
 ID AAA29338 standard; DNA; 29 BP.

```

XX AAA293338;
XX
XX 12-SEP-2000 (first entry)
XX
XX Primer V-beta-3 for MAGE-1 specific TCR-alpha chain cloning.
XX
XX human leukocyte antigen; HLA-A1; melanoma; neoplastic; MAGE-1; TCR; MHC;
XX soluble; major histocompatibility complex; antigen; T cell receptor;
XX lymphocyte; tumour; cytostatic; anti-microbial; immunosuppressive;
XX primer; alpha chain; ss.
XX
XX Homo sapiens.
XX
XX WO200031239-A1.
XX
XX 02-JUN-2000.
XX
XX 18-NOV-1999; 99WO-IL00622.
XX
XX 19-NOV-1998; 98IL-0127142.
XX
XX (YEDA ) YEDA RES & DEV CO LTD.
XX (BOLH/) BOLHUIS R L H.
XX
XX Bolhuis RLH, Eshhar Z, Willemssen RA;
XX
XX WPI; 2000-451678/39.
XX
XX Immune cells with predefined specificities useful for treating melanoma
XX and immune diseases
XX
XX Example 1; Page 29; 51pp; English.
XX
XX AAA29335-38 are oligonucleotide primers used to synthesize the alpha and
XX beta chain cDNA for a MAGE-1 specific T cell receptor from the CD8
XX positive cytotoxic T lymphocyte (CTL) clone 82/30. Novel immune cells
XX with predefined specificity, are produced by either complexing the cells
XX with an antigen-specific MHC-restricted TCR or transfecting the cells
XX with an antigen-specific MHC-restricted chimeric TCR gene. The
XX antigen-specific MHC-restricted TCR can be complexed with lymphocytic
XX cells for treatment of a tumour. Alternatively, the autologous
XX lymphocytes can be transfectd with an antigen-specific MHC-restricted
XX chimeric TCR gene encoding a single chain TCR (scFv-TCR) which binds to
XX an antigen associated with the tumour and a segment encoding a signal
XX transducing element of an immune cell. Compositions comprising the
XX immune cells may be used for the treatment of cancer (especially
XX melanomas, if the TCR binds to the MAGE-1 antigen), infectious diseases,
XX autoimmune disease and/or graft rejection.
XX
XX Sequence 29 BP; 4 A; 10 C; 11 G; 4 T; 0 other;
XX
XX
XX Query Match 61.0%; Score 12.2; DB 21; Length 29;
XX Best Local Similarity 82.4%; Pred. No. 3.4e+03;
XX Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX Qy 4 cgggctcacagtggtcg 20
XX | | | | | | | | | |
XX Db 24 CCGGCTGACAGTGCTCG 8

```

Search completed: October 2, 2001, 16:18:33  
 Job time: 15477 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:03:55 ; Search time 417.38 Seconds  
(without alignments)  
6.804 Million cell updates/sec

Title: US-09-757-100B-33

Perfect score: 15

Sequence: 1 tggcttattctcagt 15

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 460742

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCFUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	15	3	US-09-377-310-33
2	15	100.0	20	3	US-09-377-310-13
3	13	86.7	21	4	US-08-462-467B-28
4	12.4	82.7	27	1	US-08-459-568-11
5	12.4	82.7	27	2	US-08-399-411-11
6	12.4	82.7	27	3	US-08-516-859A-11
7	12.4	82.7	30	3	US-08-937-271-39
8	12	80.0	21	3	US-09-087-232A-2
9	11.8	78.7	22	4	US-09-116-498-16
10	11.8	78.7	30	1	US-07-884-811-10
11	11.8	78.7	30	1	US-07-885-971-10
12	11.8	78.7	30	1	US-08-087-783A-10
13	11.8	78.7	30	1	US-08-194-088B-10
14	11.8	78.7	30	1	US-08-435-501-9
15	11.8	78.7	30	1	US-08-435-764-9
16	11.8	78.7	30	1	US-08-792-078-9
17	11.8	78.7	30	2	US-08-194-087-10
18	11.8	78.7	30	5	PCT-US93-04648-10
19	11.8	78.7	30	5	PCT-US93-04717-9
20	11.4	76.0	21	1	US-08-271-946A-7
21	11.4	76.0	21	1	US-08-271-942A-7
22	11.4	76.0	21	3	US-08-779-916A-7
23	11.4	76.0	21	5	PCT-US95-08604-7
24	11.4	76.0	21	5	PCT-US95-08606-7
25	11.4	76.0	27	3	US-08-716-190-14
26	11.4	76.0	31	3	US-08-961-083-272
27	11.4	76.0	43	3	US-08-832-985-49

c	28	11	73.3	17	3	US-08-467-023-105	Sequence 105, Appl
	29	11	73.3	24	1	US-08-198-431-14	Sequence 14, Appl
	30	11	73.3	24	1	US-08-564-955-14	Sequence 14, Appl
	31	11	73.3	24	2	US-08-537-874-14	Sequence 14, Appl
	32	11	73.3	24	3	US-08-621-859-14	Sequence 14, Appl
	33	11	73.3	24	3	US-09-100-856A-14	Sequence 14, Appl
	34	11	73.3	24	4	US-09-075-511-14	Sequence 14, Appl
	35	11	73.3	24	4	US-09-099-015-14	Sequence 14, Appl
	36	11	73.3	31	6	5164180-1	Patent No. 5164180
	37	11	73.3	41	1	US-08-198-431-13	Sequence 13, Appl
	38	11	73.3	41	1	US-08-564-955-13	Sequence 13, Appl
	39	11	73.3	41	2	US-08-537-874-13	Sequence 13, Appl
	40	11	73.3	41	3	US-08-621-859-13	Sequence 13, Appl
	41	11	73.3	41	3	US-09-100-856A-13	Sequence 13, Appl
	42	11	73.3	41	4	US-09-075-511-13	Sequence 13, Appl
	43	11	73.3	41	4	US-09-099-015-13	Sequence 13, Appl
	44	10.8	72.0	16	2	US-08-874-266-5	Sequence 5, Appl
	45	10.8	72.0	17	1	US-08-217-016-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-377-310-33

; Sequence 33, Application US/09377310B

; Patent No. 6133031

; GENERAL INFORMATION:

; APPLICANT: Monia, Brett P.

; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase

; FILE REFERENCE: ISPH-0389

; CURRENT APPLICATION NUMBER: US/09/377,310B

; CURRENT FILING DATE: 1999-08-19

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 33

; LENGTH: 15

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: antisense sequence

US-09-377-310-33

Query Match 100.0%; Score 15; DB 3; Length 15;

Best Local Similarity 100.0%; Pred. No. 3.9;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tggcttattctcagt 15

Db 1 tggcttattctcagt 15

RESULT 2

US-09-377-310-13

; Sequence 13, Application US/09377310B

; Patent No. 6133031

; GENERAL INFORMATION:

; APPLICANT: Monia, Brett P.

; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase

; FILE REFERENCE: ISPH-0389

; CURRENT APPLICATION NUMBER: US/09/377,310B

; CURRENT FILING DATE: 1999-08-19

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 20

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-13

Query Match 100.0%; Score 15; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tggcttatcttcagt 15  
|||||  
Db 3 tggcttatcttcagt 17

RESULT 3  
US-08-462-467B-28/c  
; Sequence 28, Application US/08462467B  
; Patent No. 6210899  
; GENERAL INFORMATION:  
; APPLICANT: Rosenbaum, Jan S  
; TITLE OF INVENTION: The Use of a BMP Protein Receptor  
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells  
; TITLE OF INVENTION: Co-transfected with a Type II BMP Receptor and a Type I  
; TITLE OF INVENTION: BMP Receptor  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Procter & Gamble Company  
; STREET: 11810 East Miami River Road  
; CITY: Ross  
; STATE: OH  
; COUNTRY: USA  
; ZIP: 45061  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hersko, Bart S.  
; REGISTRATION NUMBER: 32,572  
; REFERENCE/DOCKET NUMBER: 5474R  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (513) 627-0633  
; TELEFAX: (513) 627-0260  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-462-467B-28

Query Match 86.7%; Score 13; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tggcttatcttca 13  
|||||  
Db 18 TGGCTTATCTTCA 6

RESULT 4  
US-08-459-568-11/c  
; Sequence 11, Application US/08459568  
; Patent No. 5811304  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi

; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/459,568  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,411  
; FILING DATE: 06-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 27 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-459-568-11

Query Match 82.7%; Score 12.4; DB 1; Length 27;  
Best Local Similarity 92.9%; Pred. No. 1.3e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttcag 14  
|||||  
Db 27 TGGCTTTTCTTCAG 14

RESULT 5  
US-08-399-411-11/c  
; Sequence 11, Application US/08399411  
; Patent No. 5831008  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/399,411  
; FILING DATE: 06-MAR-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:



NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1264  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-399-411-11

Query Match 82.7%; Score 12.4; DB 2; Length 27;  
Best Local Similarity 92.9%; Pred. No. 1.3e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttcag 14  
||||| |||||  
Db 27 TGGCTTTCTTCAG 14

## RESULT 6

US-08-516-859A-11/c  
Sequence 11, Application US/08516859A  
Patent No. 6069231  
GENERAL INFORMATION:  
APPLICANT: Huang, Shi  
TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/516,859A  
FILING DATE: 18-AUG-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION NUMBER: US 08/399,411  
FILING DATE: 06-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,683  
FILING DATE: 18-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1776  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-516-859A-11

Query Match 82.7%; Score 12.4; DB 3; Length 27;  
Best Local Similarity 92.9%; Pred. No. 1.3e+02;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 tggcttatcttcag 14  
||||| |||||  
Db 27 TGGCTTTCTTCAG 14

## RESULT 7

US-08-937-271-39  
Sequence 39, Application US/08937271  
Patent No. 6063386  
GENERAL INFORMATION:  
APPLICANT: Dale, James B.  
APPLICANT: Lederer, James W.  
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN  
TITLE OF INVENTION: VACCINE  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/937,271  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Rosenman, Stephen J.  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 481112.405C1  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-937-271-39

Query Match 82.7%; Score 12.4; DB 3; Length 30;  
Best Local Similarity 92.9%; Pred. No. 1.3e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttcag 14  
||||| |||||  
Db 10 TGGCTTATCTTCG 23

## RESULT 8

US-09-087-232A-2/c  
Sequence 2, Application US/09087232A  
Patent No. 6153431  
GENERAL INFORMATION:  
APPLICANT: Quillent et al.  
TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS CO-RECEPTOR  
TITLE OF INVENTION: VARIANTS ASSOCIATED WITH RESISTANCE TO VIRUS INFECTION.  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Baker & Botts, L.L.P. attn. Lisa Kole  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA

Query Match 82.7%; Score 12.4; DB 3; Length 27;  
Best Local Similarity 92.9%; Pred. No. 1.3e+02;

ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/087,232A  
FILING DATE: 28 MAY 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/048,057  
FILING DATE: 30 MAY 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: KOLE, LISA B.  
REGISTRATION NUMBER: 35,225  
REFERENCE/DOCKET NUMBER: AP 31115  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 408-2628  
TELEFAX: (212) 765-2519  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "PRIMER"  
US-09-087-232A-2

Query Match 80.0%; Score 12; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 ggcttatcttca 13  
Db 13 GGCTTATCTTCA 2  
RESULT 9  
US-09-116-498-16  
Sequence 16, Application US/09116498  
Patent No. 6251582  
GENERAL INFORMATION:  
APPLICANT: Littman, Dan R.  
Deng, Hongkui  
Unutmaz, Derya  
Ramani, Vineet N.K.  
TITLE OF INVENTION: NOVEL ALTERNATIVE G-COUPLED RECEPTORS  
ASSOCIATED WITH RETROVIRAL ENTRY INTO CELLS, METHODS OF  
IDENTIFYING THE SAME, AND DIAGNOSTIC AND THERAPEUTIC USES  
THEREOF

NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/116,498  
FILING DATE: 16-Jul-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1049-1-009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Oligonucleotides"  
HYPOTHETICAL: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 16:  
US-09-116-498-16  
Query Match 78.7%; Score 11.8; DB 4; Length 22;  
Best Local Similarity 86.7%; Pred. No. 2.7e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 tggcttatcttcagt 15  
Db 4 TGGCTTATCATCAAT 18  
RESULT 10  
US-07-884-811-10  
Sequence 10, Application US/07884811  
Patent No. 5316921  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/884,811  
FILING DATE: 19920518  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 755.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-884-811-10  
Query Match 78.7%; Score 11.8; DB 1; Length 30;  
Best Local Similarity 86.7%; Pred. No. 2.8e+02;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15  
||| |||||  
Db 7 TGACATATCTTCAGT 21

## RESULT 11

US-07-885-971-10  
; Sequence 10, Application US/07885971  
; Patent No. 5328837  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/885,971  
; FILING DATE: 19920518  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: 779  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3216  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 bases  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-07-885-971-10

Query Match 78.7%; Score 11.8; DB 1; Length 30;  
Best Local Similarity 86.7%; Pred. No. 2.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15  
||| |||||  
Db 7 TGACATATCTTCAGT 21

## RESULT 12

US-08-087-783A-10  
; Sequence 10, Application US/08087783A  
; Patent No. 5547856  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California

COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,783A  
FILING DATE: 13-Jul-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-92  
ATTORNEY/AGENT INFORMATION:  
NAME: Marschang, Diane L.  
REGISTRATION NUMBER: 35,600  
REFERENCE/DOCKET NUMBER: P0755779P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5416  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-087-783A-10

Query Match 78.7%; Score 11.8; DB 1; Length 30;  
Best Local Similarity 86.7%; Pred. No. 2.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggcttatcttcagt 15  
||| |||||  
Db 7 TGACATATCTTCAGT 21

## RESULT 13

US-08-194-088B-10  
; Sequence 10, Application US/08194088B  
; Patent No. 5580963  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/194,088B  
; FILING DATE: 09-FEB-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/884811  
; FILING DATE: 18-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gallegos, R. Thomas  
; REGISTRATION NUMBER: 32,692

REFERENCE/DOCKET NUMBER: 755D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-2614  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-194-088B-10

Query Match 78.7%; Score 11.8; DB 1; Length 30;  
Best Local Similarity 86.7%; Pred. No. 2.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggcttatcttcagt 15  
||| |||||  
Db 7 TGACATATCTTCAGT 21

RESULT 14  
US-08-435-501-9  
Sequence 9, Application US/08435501  
Patent No. 5684136  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc., Paul J. Godowski  
TITLE OF INVENTION: Receptor Activation  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,501  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/950572  
FILING DATE: 21-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 773P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-435-501-9

Query Match 78.7%; Score 11.8; DB 1; Length 30;  
Best Local Similarity 86.7%; Pred. No. 2.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggcttatcttcagt 15  
||| |||||  
Db 7 TGACATATCTTCAGT 21

RESULT 15  
US-08-435-764-9  
Sequence 9, Application US/08435764  
Patent No. 5763584  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc., Paul J. Godowski  
TITLE OF INVENTION: Receptor Activation  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,764  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/950572  
FILING DATE: 21-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 773P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-435-764-9

Query Match 78.7%; Score 11.8; DB 1; Length 30;  
Best Local Similarity 86.7%; Pred. No. 2.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggcttatcttcagt 15  
||| |||||  
Db 7 TGACATATCTTCAGT 21

Search completed: October 2, 2001, 16:03:55  
Job time: 14599 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:48 ; Search time 876.95 seconds  
(without alignments)  
10.740 Million cell updates/sec

Title: US-09-757-100B-33

Perfect score: 15

Sequence: 1 tggcttatcttcagt 15

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	15	22	AAC65565 Human focal adhesi
2	15	100.0	20	19	AAV08375 PCR primer for Cel
3	15	100.0	20	22	AAC65545 Human focal adhesi
4	13	86.7	21	17	AAT28037 PCR primer for mou
5	13	86.7	21	17	AAT27230 Antisense primer f
6	12.4	82.7	20	20	AAK94668 PCR primer used to
7	12.4	82.7	21	22	AAF95414 Human gene single
8	12.4	82.7	27	21	AAAG0107 Rat RIZ PCR primer
9	12.4	82.7	30	21	AAAS7918 Streptococcus pyog
10	12	80.0	19	22	AAF76114 CCR5/CCR2b PCR pri
11	12	80.0	21	20	AAV84128 HIV-1 co-receptor

c	12	11.8	78.7	21	20	AAV08563	Primer for Transge
	13	11.8	78.7	22	20	AAV07298	Human SIV/HIV rece
c	14	11.8	78.7	23	21	AAAG6237	Dog genomic marker
	15	11.8	78.7	30	14	AAQ52624	Primer delatK3 huH
	16	11.8	78.7	30	20	AAAI5645	Primer used to gen
	17	11.4	76.0	21	17	AAT12845	PCR 5', primer for
	18	11.4	76.0	21	17	AAT11444	Retinoblastoma gen
	19	11.4	76.0	27	16	AAT00441	N-terminal primer
	20	11.4	76.0	31	19	AAV27482	Streptococcus pneu
	21	11.4	76.0	43	20	AAV84748	Antibody heavy cha
c	22	11	73.3	17	15	AAQ55632	Juniperus Cry j i
c	23	11	73.3	20	13	AAQ29005	Sequence of PCR pr
	24	11	73.3	24	18	AAAT73032	Mutagenic PCR prim
	25	11	73.3	24	19	AAV4903	Primer used to int
	26	11	73.3	31	13	AAQ26930	Probe for the nove
	27	11	73.3	31	18	AAAT62463	G-alpha protein GP
	28	11	73.3	31	18	AAAT62485	yeast alpha-factor
	29	11	73.3	31	18	AAAT49593	Bacillus thuringie
	30	11	73.3	41	18	AAAT73031	Mutagenic PCR prim
c	31	11	73.3	41	19	AAV54902	Primer used to int
	32	11	73.3	43	18	AAAT62479	STE2 gene 3' PCR p
	33	10.8	72.0	15	19	AAV00716	ATM region b (nt 5
	34	10.8	72.0	17	20	AAAI5040	Antisense PCR prim
	35	10.8	72.0	20	13	AAQ32816	Microsatellite rep
	36	10.8	72.0	20	15	AAQ57839	Primer pair 14A W
c	37	10.8	72.0	20	16	AAQ76318	Human beta-lactogl
	38	10.8	72.0	20	19	AAV22500	Antisense oligonuc
c	39	10.8	72.0	20	19	AAV12309	Ribonucleotide red
c	40	10.8	72.0	20	20	AAQ92612	PCR primer used to
	41	10.8	72.0	20	21	AAA90730	Ribonucleotide red
c	42	10.8	72.0	20	21	AAZ88170	Mouse ssH2B PCR pr
c	43	10.8	72.0	21	20	AAAS59982	Oligonucleotide pr
c	44	10.8	72.0	22	21	AAA47518	Primer for amplify
c	45	10.8	72.0	23	14	AAQ52566	HPV-16 target sequ

#### ALIGNMENTS

RESULT 1  
AAC65565 AAC65565 standard; DNA; 15 BP.  
XX  
AC AAC65565;  
XX  
XX 12-FEB-2001 (first entry)  
DT  
XX Human focal adhesion kinase antisense sequence #31.  
DE  
XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.  
XX  
OS Homo sapiens.  
XX  
PN US6133031-A.  
XX  
PD 17-OCT-2000.  
XX  
PF 19-AUG-1999; 99US-0377310.  
XX  
PR 19-AUG-1999; 99US-0377310.  
XX  
PA (ISIS-) ISIS PHARM INC.  
XX  
PI Monia BP, Gaarde WA;  
XX  
DR WPI; 2001-006141/01.  
XX  
PT New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -

```
PS Claim 15; Column 25; 30pp; English.
XX
XX The present invention describes a number of phosphorothioate antisense
CC sequences to the human focal adhesion kinase (FAK) protein. This protein
CC is involved in integrin-mediated signal transduction, and is implicated
CC in cancer, particularly colon, breast and oral tumours, embryonic
CC development disorders, angiogenic disorders and wound healing. The
CC antisense sequences, including the one shown here, can be used in the
CC treatment of all of these.
XX
XX Sequence 15 BP; 2 A; 3 C; 3 G; 7 T; 0 other;
SQ
Query Match 100.0%; Score 15; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 tggcttatcttcagt 15
Db 1 tggcttatcttcagt 15
RESULT 2
AAV08375/c
ID AAV08375 standard; DNA; 20 BP.
XX
AC AAV08375;
XX
XX 08-FEB-1999 (first entry)
DT
XX PCR primer for Cellulose synthase coding sequence.
DE
XX Cellulose synthase; cotton; PCR primer; ss.
KW
XX Synthetic.
OS
XX Gossypium hirsutum.
XX
XX EP875575-A2.
XX
XX 04-NOV-1998.
XX
XX 31-MAR-1998; 98EP-0302489.
XX
XX 01-APR-1997; 97JP-0083133.
XX
XX (NISR) NISSHINBO IND INC.
XX
XX Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
XX WPI; 1998-559440/48.
XX
XX New DNA encoding cotton cellulose synthase proteins - useful for
XX controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
XX Example 5; Page 45; 51pp; English.
XX
XX This sequence represents a PCR primer for DNA encoding a cellulose
XX synthase of the invention. The DNA can be used for controlling cellulose
XX synthesis in a cell by introducing the DNA into the cell, expressing RNA
XX with a nucleotide sequence homologous to the DNA or a nucleotide sequence
XX complementary to the DNA. The cells are prokaryotic cells
XX i.e. acetobacterium and/or eukaryotic cells such as yeasts
XX i.e. Saccharomyces spp., plant cells (cotton plants) and cultured cells
XX of mammals etc.
XX
XX Sequence 20 BP; 10 A; 3 C; 5 G; 2 T; 0 other;
SQ
Query Match 100.0%; Score 15; DB 19; Length 20;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 tggcttatcttcagt 15
Db 1 tggcttatcttcagt 15
us-09-757-100b-33.szlim50.rng
Db 16 TGGCTTATCTTCAGT 2
RESULT 3
AAC65545
ID AAC65545 standard; DNA; 20 BP.
XX
AC AAC65545;
XX
XX 12-FEB-2001 (first entry)
DT
XX Human focal adhesion kinase antisense sequence #11.
XX
XX Human; focal adhesion kinase; FAK; signal transduction; cancer;
XX embryonic development disorder; angiogenic disorder; wound healing;
XX antisense; phosphorothioate; ss.
XX
XX Homo sapiens.
XX
XX US6133031-A.
XX
XX 17-OCT-2000.
XX
XX 19-AUG-1999; 99US-0377310.
XX
XX 19-AUG-1999; 99US-0377310.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Monia BP, Gaarde WA;
XX
XX WPI; 2001-006141/01.
XX
XX New antisense compounds for inhibiting focal adhesion kinase
XX expression, especially useful for inhibiting retinal
XX neovascularization, or for diagnosing and treating e.g. colon cancer -
XX Example 2; Column 23; 30pp; English.
XX
XX The present invention describes a number of phosphorothioate antisense
XX sequences to the human focal adhesion kinase (FAK) protein. This protein
XX is involved in integrin-mediated signal transduction, and is implicated
XX in cancer, particularly colon, breast and oral tumours, embryonic
XX development disorders, angiogenic disorders and wound healing. The
XX antisense sequences, including the one shown here, can be used in the
XX treatment of all of these.
XX
XX Sequence 20 BP; 5 A; 3 C; 4 G; 8 T; 0 other;
SQ
Query Match 100.0%; Score 15; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 tggcttatcttcagt 15
Db 3 tggcttatcttcagt 17
RESULT 4
AAT28037/c
ID AAT28037 standard; DNA; 21 BP.
XX
AC AAT28037;
XX
XX 31-DEC-1996 (first entry)
DT
XX PCR primer for mouse BRK-3 cDNA amplification.
XX
XX BMP type II receptor kinase-3; BRK-3; bone morphogenetic protein;
XX primer; PCR; polymerase chain reaction; ss.
XX
```



OS Synthetic.  
 XX WO9614579-A1.  
 XX 17-MAY-1996.  
 PD  
 XX 30-OCT-1995; 95WO-US14027.  
 XX 05-JUN-1995; 95US-0462467.  
 PR 04-NOV-1994; 94US-0334178.  
 XX (PROC ) PROCTER & GAMBLE CO.  
 PA  
 XX Rosenbaum JS;  
 PI WPI; 1996-251887/25.  
 DR  
 XX Assays for bone morphogenetic protein activities - using complex of  
 PT BMP type I receptor kinase protein and BMP receptor kinase protein  
 PT BRK-3  
 XX  
 XX Example 9; Page 32; 101pp; English.  
 XX A PCR primer (AAT28037) includes the reverse complement of a  
 CC partial cDNA clone isolated from NIH3T3 mouse embryonic  
 CC fibroblasts. It was used with an anchor primer (AAT28038) to  
 CC amplify the 5' end of mouse bone morphogenetic protein (BMP) type  
 CC receptor kinase protein-3 (BRK-3) cDNA. Following further PCR  
 CC amplification and cloning into pUT6, a cDNA clone (AAT28020) coding  
 CC for full-length murine BRK-3 (AA95224) was obt'd. The PCR primer is  
 XX the reverse complement of nucleotides 968-948 of this clone.  
 XX  
 SQ Sequence 21 BP; 6 A; 4 C; 6 G; 5 T; 0 other;

Query Match 86.7%; Score 13; DB 17; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tggcttatcttca 13  
 |||||  
 Db 18 TGGCTTATCTTCA 6

RESULT 5  
 AAT27230/C  
 ID AAT27230 standard; DNA; 21 BP.  
 AC  
 XX AAT27230;  
 XX  
 XX 23-AUG-1996 (first entry)  
 DT  
 DE  
 XX Antisense primer for mouse BRK-3 gene 5'-end amplification.  
 XX  
 XX Primer; antisense; 5'-RACE; PCR; polymerase chain reaction;  
 KW mouse; bone morphogenetic protein receptor kinase-3; antibody;  
 KW diagnostic; bone disorder; osteogenic; reporter gene;  
 KW bone morphogenetic protein-agonist; drug screening;  
 KW bone morphogenetic protein-antagonist; hormone-responsive element; ss.  
 OS  
 XX Synthetic.  
 XX  
 XX WO9614412-A2.  
 XX  
 XX 17-MAY-1996.  
 PD  
 XX 30-OCT-1995; 95WO-US14085.  
 XX 04-NOV-1994; 94US-0334179.  
 PR  
 XX (PROC ) PROCTER & GAMBLE CO.  
 PA  
 XX Nohno T, Rosenbaum JS;  
 PI

XX WPI; 1996-251762/25.  
 DR  
 XX Isolated bone morphogenic protein receptor kinase protein - used to  
 PT determine if a test cpl. is capable of binding to, or is  
 PT (ant)agonist of BMP receptor kinase protein transcription  
 XX  
 XX Example 9; Page 30; 87pp; English.  
 XX  
 XX This primer corresponds to the sequence of plasmid pSPORT1/N89-5,  
 CC and is the reverse complement of nucleotides 968-948 in a gene  
 CC encoding mouse bone morphogenetic protein (BMP) receptor type-II  
 CC kinase-3 (BRK-3), which induces cellular differentiation in response  
 CC to BMP. The primer may be used with anchor primer AAT27231 in 5'-rapid  
 CC amplification of cDNA ends (5'-RACE), to isolate 600 bp at the 5'-end  
 CC of the BRK-3 gene, by PCR amplification of NIH3T3 mouse embryo  
 CC fibroblast (ATCC CRJ 1658) cDNA. To reduce background, a 2nd round  
 CC of PCR is performed using nested primer AAT27232 with the anchor  
 CC primer. This results in isolation of 600 bp from the 5'-end of the  
 CC gene, which may be used to assemble a full-length clone (AAT27225) by  
 CC PCR using primers AAT27223-4 and cloning in phage Bluescript-II-SK(-)  
 CC in Escherichia coli DH5-alpha. The receptor and antibodies against  
 CC it may be used in diagnostic assays for BMP disorders, or in  
 CC therapy to bind or scavenge BMPs. In addition, expression of the  
 CC BRK-3 gene along with a reporter gene under the control of a  
 CC hormone-responsive element in a cell culture may be used to screen  
 CC compounds for BRK-agonist or -antagonist activity, by monitoring  
 CC reporter gene expression.  
 XX  
 SQ Sequence 21 BP; 6 A; 4 C; 6 G; 5 T; 0 other;

Query Match 86.7%; Score 13; DB 17; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tggcttatcttca 13  
 |||||  
 Db 18 TGGCTTATCTTCA 6

RESULT 6  
 AAX94668  
 ID AAX94668 standard; DNA; 20 BP.  
 AC  
 XX AAX94668;  
 XX  
 XX 13-SEP-1999 (first entry)  
 DT  
 DE  
 XX PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
 XX  
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; PCR primer; ss.  
 XX  
 XX Synthetic.  
 OS  
 OS Chlamydia pneumoniae.  
 XX  
 XX WO9927105-A2.  
 PN  
 XX 03-JUN-1999.  
 PD  
 XX 20-NOV-1998; 98WO-IB01890.  
 PF  
 XX 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 XX (GEST ) GENSET.  
 PA  
 XX Griffais R;  
 PI  
 XX WPI; 1999-357842/30.  
 DR  
 XX



CC nucleus. RIZ gene mutations may be implicated in various cancers such as  
 CC melanoma, neuroblastoma, leukaemia and breast cancer, and so the RIZ  
 CC gene may be used in gene therapy for these disorders. Since RIZ protein  
 CC is implicated in cell cycle arrest, inhibition of RIZ activity may be  
 CC useful in neurodegenerative disorder therapy e.g. for Parkinson's,  
 CC Huntington's or Alzheimer's disease, paralysis or motor neurone  
 CC disorders, or cardiac disorders e.g. heart disease, where the ability to  
 CC induce neural/ cardiac tissue proliferation would be useful. The present  
 CC sequence was used in the construction of vectors which were used to clone  
 CC mammalian RIZ coding sequences.

XX SQ Sequence 27 BP; 9 A; 7 C; 9 G; 2 T; 0 other;

Query Match 82.7%; Score 12.4; DB 21; Length 27;

Best Local Similarity 92.9%; Pred. No. 6.2e+02;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttcag 14

Db 27 TGGCTTTCTTCAG 14

RESULT 9

AAA57918

ID AAA57918 standard; DNA; 30 BP.

XX AC AAA57918;

XX DT 10-OCT-2000 (first entry)

XX DE Streptococcus pyogenes M19 gene PCR primer, SEQ ID NO:39.

XX KW Multivalent hybrid M protein: group A streptococcus; serotype;

XX KW immunogenic; sero-specific antibody; streptococcal infection;

XX KW cross reactivity; vaccine; acute rheumatic fever; ARF;

XX KW rheumatic heart disease; streptococcal pharyngitis; strep throat;

XX KW pneumonia; PCR primer; ss.

XX OS Streptococcus pyogenes.

XX PN US6063386-A.

XX PD 16-MAY-2000.

XX PF 15-SEP-1997; 97US-0937271.

XX PR 16-SEP-1992; 92US-0945954.

XX PA (UYTE-) UNIV TENNESSEE RES CORP.

XX PI Lederer JW, Dale JB;

XX DR WPI; 2000-364475/31.

XX PT New immunogenic recombinant hybrid M protein comprising amino-terminal

XX PT peptide fragments of streptococcal M protein useful as vaccine against

XX PT rheumatic fever and infections leading to rheumatic fever -

XX PS Disclosure; Column 13-14; 62pp; English.

XX CC The invention relates to multivalent immunogenic hybrid group A

XX CC streptococcal M proteins comprising N-terminal peptide fragments of M

XX CC proteins that elicit opsonic antibodies against multiple serotypes

XX CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies

XX CC generated using the hybrid proteins are against one or more M protein

XX CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.

XX CC The invention also encompasses a recombinant DNA molecule comprising a

XX CC nucleotide sequence that encodes a multivalent hybrid M protein; and a

XX CC method for immunising a mammal against streptococcal infections,

XX CC comprising administering an immunogenic multivalent hybrid M protein to

XX CC the mammal. The multivalent hybrid M proteins are useful for eliciting

XX CC opsonic or protective antibodies to the M proteins of different

CC serotypes of group A streptococci, and may therefore be used as vaccines  
 CC to protect against and control infection by type A streptococci. Type A  
 CC streptococci are not only responsible for streptococcal pharyngitis  
 CC (strep throat), forms of pneumonia and a condition resembling toxic  
 CC shock, but are also involved in the development of acute rheumatic fever  
 CC (ARF) and rheumatic heart disease. In a patient with ARF, antibodies  
 CC formed during a group A streptococcal infection are also cross-reactive  
 CC with heart tissue, which indicates that the streptococci and host tissue  
 CC contain similar antigenic motifs. The new multivalent vaccines are  
 CC capable of raising sero-specific antibodies against various serotypes of  
 CC group A streptococci which are not cross-reactive with human heart  
 CC tissue. Sequences AAA57903-A57904 and AAA57909-A57919 represent PCR  
 CC primers used in the generation of DNA (AAA57893-A57902) encoding  
 CC multivalent hybrid Streptococcus pyogenes M proteins (AAB03113-B03117,  
 CC AAB03119-B03121, AAB03123-B03124).

XX SQ Sequence 30 BP; 3 A; 6 C; 12 G; 9 T; 0 other;

Query Match 82.7%; Score 12.4; DB 21; Length 30;

Best Local Similarity 92.9%; Pred. No. 6.3e+02;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tggcttatcttcag 14

Db 10 tggcttatcttcag 23

RESULT 10

AAF76114/C

ID AAF76114 standard; DNA; 19 BP.

XX AC AAF76114;

XX DT 22-MAY-2001 (first entry)

XX DE CCR5/CCR2b PCR primer, SEQ ID:18, used to genotype HIV susceptibility.

XX KW CC chemokine receptor; beta chemokine receptor; CCR; human;

XX KW CCR5; CCR2; polymorphism; genotyping; HIV-1 transmission;

XX KW infection susceptibility; AIDS; acquired immunodeficiency syndrome;

XX KW disease progression; chromosome 3p21-22; PCR primer; ss.

XX OS Homo sapiens.

XX PN WO200112857-A2.

XX PD 22-FEB-2001.

XX PF 11-AUG-2000; 2000WO-US22255.

XX PR 12-AUG-1999; 99US-0148530.

XX PA (UABR-) UAB RES FOUND.

XX PI Tang J, Kaslow RA;

XX DR WPI; 2001-211235/21.

XX PT Surveying CC beta chemokine receptor (CCR) genotypes in population,

XX PT involves amplifying genomic DNA of individuals with experimental and

XX PT control primer combinations, size-separating amplicons and determining

XX PS CCR genotype -

XX CC Claim 1; Page 42; 118pp; English.

XX CC The invention relates to a method of surveying the CC (beta) chemokine

XX CC receptor (CCR) genotypes in a population. The method is particularly

XX CC applied to the human CCR5 and CCR2 genes located on chromosome 3p21-22,

XX CC which encode co-receptors for HIV-1. The method involves obtaining

XX CC genomic DNA samples from a representative number of individuals within

XX CC a population; combining each sample with experimental and control primer

XX CC combinations to produce primer-annealed DNA; amplifying the DNA to

CC produce amplicons; separating the amplicons by size; determining the CCR  
 CC genotype based upon the presence of CCR alleles; and compiling the  
 CC genotypes determined. The method is particularly applied to the human  
 CC CCR5 and CCR2 genes, which encode co-receptors for HIV-1. Polymorphisms  
 CC in these genes are associated with a variation in the susceptibility of  
 CC an individual to infection by HIV-1, or with a variation in the disease  
 CC progression of AIDS after infection. The invention specifically claims  
 CC the experimental PCR primers AAF76098-AAF76112, and the control PCR  
 CC primers AAF76113-AAF76114 for surveying CCR5 and CCR2b genotypes.  
 CC The method of the invention fulfills a longstanding need for the  
 CC development of a rapid and informative genotyping strategy that can be  
 CC readily applied to analyse CCR5, CCR2 and related genetic variants,  
 CC and to evaluate the relationship of each genotype to HIV transmission  
 CC and disease progression. The present sequence represents a human  
 CC CCR5/CCR2b control PCR primer for use in the method of the  
 CC invention.  
 CC  
 CC Sequence 19 BP; 6 A; 4 C; 5 G; 4 T; 0 other;

Query Match 80.0%; Score 12; DB 22; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ggcttatcttca 13  
 |||||  
 Db 15 GGCTTATCTTCA 4

RESULT 11  
 AAV84128/c  
 ID AAV84128 standard; DNA; 21 BP.  
 XX  
 AC AAV84128;  
 XX  
 DT 15-MAR-1999 (first entry)  
 XX  
 DE HIV-1 co-receptor CCR5-delta32 variant PCR primer delta-32 inv.  
 XX  
 KW HIV-1; CCR5; CCR5-delta32; co-receptor; infection; diagnosis; AIDS;  
 KW gene therapy; human; PCR; primer; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO9854317-A1.  
 XX  
 PD 03-DEC-1998.  
 XX  
 PF 29-MAY-1998; 98WO-EP03437.  
 XX  
 PR 30-MAY-1997; 97US-0048057.  
 XX  
 PA (MOND-) FOND MONDIALE RECH & PREVENTION SIDA.  
 XX  
 PI Arenzana Siesdedos F, Beretta A, Braun J, Quillient C;  
 XX  
 DR WPI; 1999-059835/05.  
 XX  
 PT New CCR5 variant protein of the HIV-1 co-receptor - useful in  
 PT developing resistance of CCR5-expressing cells to HIV-1 infection  
 XX  
 PS Claim 7; Page 8; 55pp; English.  
 XX  
 CC This is the nucleotide sequence of a claimed DNA primer, designated  
 CC delta-32 inv, which can be used for the amplification of HIV-1  
 CC CCR5 co-receptor nucleic acids. It corresponds to nucleotide  
 CC 955 of wild-type CCR5 cDNA (see AAV84126). 4 pairs of primers (see  
 CC AAV84127-34) are provided. These can be used in claimed methods for  
 CC detecting and identifying the presence of a CCR5 variant in a  
 CC sample. Primer pair delta-32 inv and delta-32 sens (see AAV84127)  
 CC are particularly useful in detecting the presence of variant  
 CC CCR5-delta32 (see AAV84159) that contains a 32 bp deletion when

CC compared to the wild-type allele. Individuals homozygous for the  
 CC variant allele show resistance to HIV-1 infection. The detection  
 CC of CCR5 variants may be used to identify individuals at lower risk  
 CC of infection relative to the general population who, if infected,  
 CC may exhibit slower progression to AIDS.  
 XX  
 SQ Sequence 21 BP; 7 A; 6 C; 5 G; 3 T; 0 other;

Query Match 80.0%; Score 12; DB 20; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 1e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ggcttatcttca 13  
 |||||  
 Db 13 GGCTTATCTTCA 2

RESULT 12  
 AAV08563/c  
 ID AAV08563 standard; DNA; 21 BP.  
 XX  
 AC AAV08563;  
 XX  
 DT 12-FEB-1999 (first entry)  
 XX  
 DE Primer for Transgene for epitope tagged TBP protein.  
 XX  
 KW TATA-box binding protein; epitope-tagged TBP; transcription complex; TAF;  
 KW TBP associated factor; TAF-interaction factor; gene expression regulator;  
 KW PCR primer; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN EP881288-A1.  
 XX  
 PD 02-DEC-1998.  
 XX  
 PF 26-MAY-1998; 98EP-0109516.  
 XX  
 PR 26-MAY-1997; 97EP-0108433.  
 XX  
 PA (FARH ) HOECHST AG.  
 XX  
 PI Berglund E, Kirschbaum B, Meisterernst M, Polites G;  
 XX  
 DR WPI; 1999-001394/01.  
 XX

Transgenic animal expressing epitope-tagged TATA-box binding protein  
 - for isolating higher-order transcription complexes and specific  
 factors that associate with the protein, useful as potential  
 therapeutic agents

Example 4; Page 11; 38pp; English.

CC This sequence is a PCR primer for DNA encoding an epitope-tagged TATA-box  
 CC binding protein (TBP) that is expressed by the transgenic non-human  
 CC animals of the invention. The animals are used to produce TBP. TBP is  
 CC used to isolate and characterise higher-order transcription complexes  
 CC (from different tissue and cell types, optionally at different  
 CC developmental stages). It is also used to identify new and/or specific  
 CC TBP associated factors (TAFs, e.g. transcription factors, activators or  
 CC inhibitors) and TAF-interaction factors, and to raise antibodies against  
 CC TBP. The TAFs may be useful for regulating gene expression,  
 CC e.g. disease-related genes, so are potential pharmaceuticals, also for  
 CC identifying human analogues for use in drug screening. The antibodies are  
 CC used for affinity purification of TBP and its complexes. TBP can isolate  
 CC transcription complexes from a wide variety of different tissues and  
 CC cells (contrast known methods that are limited to isolation from a  
 CC particular cell type).

Sequence 21 BP; 6 A; 4 C; 8 G; 3 T; 0 other;

Query Match 78.7%; Score 11.8; DB 20; Length 21;  
Best Local Similarity 86.7%; Pred. No. 1.3e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tggcttatcttcagt 15.  
|||||  
DB 19 TGGCTTAACTTCACT 5

RESULT 13  
AAQ52624  
ID AAX07298 standard; DNA: 22 BP.

AC AAX07298;

XX 21-MAY-1999 (first entry)

DE Human SIV/HIV receptor Bob downstream PCR primer.

XX SIV receptor; HIV receptor; Bob; translocation promoting agent;  
KW Bonzo; G protein coupled receptor; Chemokine receptor; AIDS;  
KW Infection; virus transmission; PCR; primer; ss.

XX Synthetic.

OS Homo sapiens.

XX WO9903888-A1.

PN 28-JAN-1999.

XX 17-JUL-1998; 98WO-US14857.

XX 17-JUL-1997; 97US-0896155.

XX (UUNY ) UNIV NEW YORK STATE.

XX Deng H, Kewalramani VN, Littman DR, Unutmaz D;

XX WPI; 1999-132164/11.

XX New nucleic acid encoding the human translocation promoter Bonzo -  
PT used to screen for potential agents for treating acquired immune  
PT deficiency syndrome

XX Example 1; Page 51; 97pp; English.

CC This primer corresponds to a region 550 bp downstream of the ATG  
CC initiation codon of human Bob (see AAX07292), a new HIV/SIV  
CC traslocation promoter agent which acts with CD4 to serve as a  
CC receptor for retroviral entry. It was used with an upstream  
CC primer (see AAX07297) in the RT-PCR amplification of lymphocyte  
CC RNA. Bob mRNA was detected in PHA-stimulated peripheral blood  
CC mononuclear cells (PBMC), purified T cells, and weakly in  
CC unstimulated PBMC. Expression was either minimal or absent in  
CC macaques and B cells. Human, African green monkey and pig-tailed  
CC macaque genes (see AAX07289-94) encoding 2 receptors, termed Bonzo  
CC and Bob (see AAX97783-88), that are associated with retroviral entry  
CC into cells have been identified. Diagnostic and therapeutic uses  
CC of these novel translocation promoting agents are provided,  
CC including a method for identifying potential agents for treatment  
CC of SIV- and HIV-induced AIDS.

XX Sequence 22 BP; 6 A; 5 C; 4 G; 7 T; 0 other;

Query Match 78.7%; Score 11.8; DB 20; Length 22;  
Best Local Similarity 86.7%; Pred. No. 1.3e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tggcttatcttcagt 15  
|||||||

DB 4 tggcttatcatcaat 18

RESULT 14

AAA66237/c

ID AAA66237 standard; DNA: 23 BP.

XX AAA66237;

XX 09-OCT-2000 (first entry)

XX Dog genomic marker oligonucleotide sequence SEQ ID NO:99.

XX Dog; genome; genomic marker; radiation hybrid map; identification;  
KW chromosome location; gene marker; polymorphic microsatellite marker;  
KW phenotype; behaviour; pedigree; ss.

XX Canis familiaris.

XX WO200029615-A2.

XX 25-MAY-2000.

XX 15-NOV-1999; 99WO-IB01907.

XX 13-NOV-1998; 98US-0108193.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Galibert F, Andre C;

XX WPI; 2000-387821/33.

XX New radiation hybrid map of the dog, Canine familiaris, genome, useful  
PT for e.g. identifying genes implicated in phenotypic and behavioral  
PT traits or in genetic diseases and for studying dog pedigrees -

XX Claim 1; Page 57; 87pp; English.

CC The present invention describes a radiation hybrid map of the dog  
CC (Canine familiaris) genome comprising the genome location of a marker  
CC selected from AAA66139 to AAA66942. The radiation hybrid map is useful  
CC for identifying and localising dog genes, since it covers approximately  
CC 80 % of the dog genome and provides a dense map integrating different  
CC types (i.e. Type I and Type II) of markers. The map and the dog genome  
CC markers (or complementary sequences) are especially useful to identify  
CC genes responsible for phenotypic and behavioural traits in dogs, to  
CC identify morbid genes, to analyse diseases and identify implicated genes  
CC in such diseases and their alleles, and to study dog pedigrees. They  
CC may also be useful for isolating corresponding human gene sequences  
CC e.g. genes involved in genetic diseases.

XX Sequence 23 BP; 10 A; 6 C; 5 G; 2 T; 0 other;

Query Match 78.7%; Score 11.8; DB 21; Length 23;  
Best Local Similarity 86.7%; Pred. No. 1.3e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tggcttatcttcagt 15

|||||  
DB 17 TGGCTTGTTCAGT 3

RESULT 15

AAQ52624

ID AAX052624 standard; DNA: 30 BP.

AC AAX052624;

XX 03-JUN-1994 (first entry)

XX Primer deltaK3 huHGF.

XX Polymerase chain reaction; PCR; amplify; primer; PRK 5.1 huHGF;  
KW human hepatocyte growth factor; huHGF; variant; E. coli; human;  
KW fetal; kidney 293 cells; hepatocyte; ligand; receptor; activation; ss.  
XX OS Synthetic.  
XX PN WO9323550-A.  
XX PD 25-NOV-1993.  
XX PF 17-MAY-1993; 93WO-US04717.  
XX PR 18-MAY-1992; 92US-0884811.  
PR 18-MAY-1992; 92US-0885971.  
XX PR 22-SEP-1992; 92US-0950572.  
XX PA (GETH ) GENENTECH INC.  
XX PI Godowski PJ;  
XX DR WPI; 1993-386581/48.  
XX PT Activation of receptors, partic. of growth factors - using  
PT conjugate of two ligands capable of binding to receptors, for  
PT mimicking biological activity  
XX Example 1; Page 49; 103pp; English.  
XX CC The sequences given in AQ52616-28 are primers which were used in the  
CC generation of PRK 5.1 human hepatocyte growth factor (huHGF) variant  
CC molecules. The amplification products from these reactions were  
CC cloned in E. coli cells and were then used to transfect human fetal  
CC kidney 293 cells. The expressed proteins were used to study the  
CC ability of huHGF to induce DNA synthesis of hepatocytes in primary  
CC culture. The expressed proteins may be used as first and/or second  
CC ligands in a conjugate for receptor activation. These conjugates  
CC may be used to activate the respective receptors and thereby mimic  
CC the biological activity of the corresponding native ligand.  
XX SQ Sequence 30 BP; 7 A; 7 C; 4 G; 12 T; 0 other;

Query Match 78.7%; Score 11.8; DB 14; Length 30;  
Best Local Similarity 86.7%; Pred. NO. 1.3e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tggcttatcttcagt 15  
||| |||||  
Db 7 tgacatatcttcagt 21

Search completed: October 2, 2001, 16:18:49  
Job time: 15493 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:03:54 ; Search time 417.38 Seconds  
(without alignments)  
6.804 Million cell updates/sec

Title: US-09-757-100B-31

Perfect score: 15  
Sequence: 1 taccagggtgagtct 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 460742

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	15	100.0	15	3	US-09-377-310-31
2	15	100.0	20	3	US-09-377-310-11
c 3	12.4	82.7	21	1	US-08-136-118-7
c 4	12.4	82.7	22	2	US-08-117-952-668
5	11.8	78.7	20	4	US-09-101-886B-71
6	11.8	78.7	34	2	US-08-988-128-15
c 7	11.8	78.7	48	2	US-08-865-675-6
c 8	11.8	78.7	48	2	US-09-237-510-6
c 9	11.4	76.0	20	4	US-09-488-671-119
10	11.4	76.0	30	1	US-08-123-702-34
11	11	73.3	20	2	US-09-015-876-20
12	11	73.3	20	2	US-08-609-443B-53
c 13	11	73.3	21	4	US-09-273-565-86
14	11	73.3	28	1	US-08-456-103-4
15	11	73.3	28	2	US-09-067-908-4
16	11	73.3	45	1	US-09-015-876-21
17	10.8	72.0	20	3	US-09-257-799-49
18	10.8	72.0	20	3	US-08-920-919A-49
c 19	10.8	72.0	21	1	US-08-477-877B-78
20	10.8	72.0	21	1	US-08-477-877B-79
c 21	10.8	72.0	21	1	US-08-472-281A-78
22	10.8	72.0	21	1	US-08-472-281A-79
c 23	10.8	72.0	21	2	US-08-477-989B-78
24	10.8	72.0	21	2	US-08-477-989B-79
25	10.8	72.0	22	1	US-07-916-034-6
c 26	10.8	72.0	26	1	US-08-482-882-87
c 27	10.8	72.0	26	1	US-08-483-389-87

c 28	10.8	72.0	26	2	US-08-487-113D-87	Sequence 87, Appl
c 29	10.8	72.0	26	2	US-08-473-503-87	Sequence 87, Appl
c 30	10.8	72.0	26	2	US-08-483-932-87	Sequence 87, Appl
c 31	10.8	72.0	26	2	US-08-720-420A-87	Sequence 87, Appl
c 32	10.8	72.0	26	3	US-08-714-017-87	Sequence 87, Appl
c 33	10.8	72.0	26	3	US-08-475-680-87	Sequence 87, Appl
c 34	10.8	72.0	29	1	US-08-530-492-69	Sequence 69, Appl
c 35	10.8	72.0	29	4	US-08-906-517-69	Sequence 69, Appl
c 36	10.8	72.0	39	4	US-07-987-264-44	Sequence 44, Appl
c 37	10.8	72.0	40	1	US-08-482-882-100	Sequence 100, App
c 38	10.8	72.0	40	1	US-08-483-389-100	Sequence 100, App
c 39	10.8	72.0	40	2	US-08-487-113D-100	Sequence 100, App
c 40	10.8	72.0	40	2	US-08-473-503-100	Sequence 100, App
c 41	10.8	72.0	40	2	US-08-483-932-100	Sequence 100, App
c 42	10.8	72.0	40	2	US-08-720-420A-100	Sequence 100, App
c 43	10.8	72.0	40	3	US-08-714-017-100	Sequence 100, App
c 44	10.8	72.0	40	3	US-08-475-680-100	Sequence 100, App
c 45	10.8	72.0	42	1	US-08-482-882-68	Sequence 68, Appl

## ALIGNMENTS

RESULT 1  
US-09-377-310-31  
: Sequence 31, Application US/09377310B  
: Patent No. 6133031  
: GENERAL INFORMATION:  
: APPLICANT: Monia, Brett P.  
: TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
: FILE REFERENCE: ISPH-0389  
: CURRENT APPLICATION NUMBER: US/09/377,310B  
: CURRENT FILING DATE: 1999-08-19  
: NUMBER OF SEQ ID NOS: 43  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 31  
: LENGTH: 15  
: TYPE: DNA  
: ORGANISM: Artificial Sequence  
: FEATURE:  
: OTHER INFORMATION: antisense sequence  
US-09-377-310-31

Query Match 100.0%; Score 15; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. NO. 4.1;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taccagggtgagtct 15  
| | | | | | | | | | | | | | | |  
Db 1 taccagggtgagtct 15

RESULT 2  
US-09-377-310-11  
: Sequence 11, Application US/09377310B  
: Patent No. 6133031  
: GENERAL INFORMATION:  
: APPLICANT: Monia, Brett P.  
: TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
: FILE REFERENCE: ISPH-0389  
: CURRENT APPLICATION NUMBER: US/09/377,310B  
: CURRENT FILING DATE: 1999-08-19  
: NUMBER OF SEQ ID NOS: 43  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 11  
: LENGTH: 20  
: TYPE: DNA  
: ORGANISM: Artificial Sequence

;  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-11

Query Match 100.0%; Score 15; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taccagggtgagtct 15  
| | | | | | | | | | | | | | | | | | | | | |  
Db 3 taccagggtgagtct 17

RESULT 3  
US-08-136-118-7/c  
; Sequence 7, Application US/08136118  
; Patent No. 5380969  
; GENERAL INFORMATION:  
; APPLICANT: HOKE, Glenn D  
; APPLICANT: BRADLEY, Matthews O  
; APPLICANT: WILLIAMS, Taffy J  
; APPLICANT: LEE, Che-Hung  
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES DIRECTED  
; AGAINST HUMAN ICAM-1  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Naval Medical Res. & Dev. Cmd.  
; STREET: 8901 Wisconsin Ave.  
; CITY: Bethesda  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20889-5606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/136.118  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/918,259  
; FILING DATE: 24-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, A. David  
; REGISTRATION NUMBER: 24,743  
; REFERENCE/DOCKET NUMBER: N.C. 75,776  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 295-6759  
; TELEFAX: (202) 295-1022  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; HYPOTHEICAL: NO  
; ANTI-SENSE: YES  
US-08-136-118-7

Query Match 82.7%; Score 12.4; DB 1; Length 21;  
Best Local Similarity 92.9%; Pred. No. 1.3e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 taccagggtgagtc 14  
| | | | | | | | | | | | | | | | | | | | | |  
Db 18 TTCCAGGTGAGTC 5

RESULT 4

US-08-117-952-668/c  
; Sequence 668, Application US/08117952  
; Patent No. 5851760  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen A.  
; APPLICANT: Smith, Michael W.  
; TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE  
; SAMPLED MAPS OF COMPLEX GENOMES  
; NUMBER OF SEQUENCES: 797  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/117,952  
; FILING DATE: 07-SEP-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/078,471  
; FILING DATE: 15-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: P41 9423  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4737  
; TELEFAX: 619-546-9392  
; INFORMATION FOR SEQ ID NO: 668:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Oligonucleotide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-117-952-668

Query Match 82.7%; Score 12.4; DB 2; Length 22;  
Best Local Similarity 92.9%; Pred. No. 1.3e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 acccagggtgagtct 15  
| | | | | | | | | | | | | | | | | | | | | |  
Db 14 ATCCAGGTGAGTCT 1

RESULT 5  
US-09-101-886B-71  
; Sequence 71, Application US/09101886B  
; Patent No. 6197507  
; GENERAL INFORMATION:  
; APPLICANT: BERG, THOMAS  
; APPLICANT: TOLLERSRUD, OLE K  
; APPLICANT: NILSEN, OIVIND  
; TITLE OF INVENTION: GENETIC TEST FOR ALPHA-MANNOSIDOSIS  
; NUMBER OF SEQUENCES: 104  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BARBARA G. ERNST  
; STREET: 555 13TH STREET, NW SUITE 701E  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20004



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/101,886B  
FILING DATE: 29-JANUARY-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB97/00109  
FILING DATE: 12-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: ERNST, BARBARA G  
REGISTRATION NUMBER: 30,377  
REFERENCE/DOCKET NUMBER: 1181-240  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-783-6040  
TELEFAX: 202-783-6031  
INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-101-886B-71

Query Match 78.7%; Score 11.8; DB 4; Length 20;  
Best Local Similarity 86.7%; Pred. No. 2.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taccacagtgagtct 15  
| ||||| ||||| |  
Db 4 TGCCACAGTGAGTGT 18

RESULT 6  
US-08-988-128-15  
Sequence 15, Application US/08988128  
Patent No. 5994505  
GENERAL INFORMATION:  
APPLICANT: Ting, Jenny Pan-Yung  
APPLICANT: Chin, Keh-Chin  
TITLE OF INVENTION: No. 5994505el Forms of Class II MHC  
TITLE OF INVENTION: Transactivator (CIITA)  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Myers Bigel Sibley & Sajovec, P.A.  
STREET: P.O. Box 37428  
CITY: Raleigh  
STATE: No. 5994505th Carolina  
COUNTRY: USA  
ZIP: 27627  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/988,128  
FILING DATE: 10-DEC-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Biswas, Sorojini J.  
REGISTRATION NUMBER: 39,111  
REFERENCE/DOCKET NUMBER: 5470-136  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 854-1400  
TELEFAX: (919) 854-1401

INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-988-128-15

Query Match 78.7%; Score 11.8; DB 2; Length 34;  
Best Local Similarity 86.7%; Pred. No. 2.9e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taccacagtgagtct 15  
| ||||| || |||||  
Db 16 TACCACAGCTGTGTCT 30

RESULT 7  
US-08-865-675-6/c  
Sequence 6, Application US/08865675  
Patent No. 5928869  
GENERAL INFORMATION:  
APPLICANT: Nadeau, James G.  
APPLICANT: Pitner, James B.  
APPLICANT: Linn, Carl P.  
APPLICANT: Schram, James L.  
TITLE OF INVENTION: DETECTION OF NUCLEIC ACIDS BY  
TITLE OF INVENTION: FLUORESCENCE QUENCHING  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: R. J. Rodrick, Becton Dickinson and Company  
STREET: 1 Becton Drive  
CITY: Franklin Lakes  
STATE: NJ  
COUNTRY: US  
ZIP: 07417  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/865,675  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fugitt, Donna R.  
REGISTRATION NUMBER: 32,135  
REFERENCE/DOCKET NUMBER: P-3746  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-865-675-6

Query Match 78.7%; Score 11.8; DB 2; Length 48;  
Best Local Similarity 86.7%; Pred. No. 3e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taccacagtgagtct 15  
| || |||| |||||  
Db 41 TACTCAGATGAGTCT 27

RESULT 8  
US-09-237-510-6/c  
Sequence 6, Application US/09237510  
Patent No. 5958700

```

: GENERAL INFORMATION:
: APPLICANT: Nadeau, James G.
: APPLICANT: Pitner, James B.
: APPLICANT: Linn, Carl P.
: APPLICANT: Schram, James L.
: TITLE OF INVENTION: DETECTION OF NUCLEIC ACIDS BY
: TITLE OF INVENTION: FLUORESCENCE QUENCHING
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: R. J. Rodrick, Becton Dickinson and Company
: STREET: 1 Becton Drive
: CITY: Franklin Lakes
: STATE: NJ
: COUNTRY: US
: ZIP: 07417
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/237,510
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Fugit, Donna R.
: REGISTRATION NUMBER: 32,135
: REFERENCE/DOCKET NUMBER: P-3746
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 48 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-237-510-6

```

Query Match	78.7%	Score 11.8;	DB 2;	Length 48;
Best Local Similarity	86.7%	Pred. No. 3e+02;		
Matches 13; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	taccacaggtgagtct	15	
Db	41	TACTCAGATGAGTCT	27	

```

RESULT          9
US-09-488-671-119/c
; Sequence 119, Application US/09488671A
; Patent No. 6187545
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Madeline M. Butler
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Lex W. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF PEPCK-CYTOSOLIC EXPRESSION
; FILE REFERENCE: RTS-0123
; CURRENT APPLICATION NUMBER: US/09/488,671A
; CURRENT FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 177
; SEQ ID NO 119
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-09-488-671-119

```

```
Query Match          76.0%; Score 11.4; DB 4; Length 20;
Best Local Similarity 92.3%; Pred. No. 4.8e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY      3  cccaggtgagtct 15
      || |||||
Db      15  CCAAGGTGAGTCT 3

RESULT 10
US-08-123-702-34
; Sequence 34, Application US/08123702
; Patent No. 5604131
; GENERAL INFORMATION:
; APPLICANT: Wadsworth, Samuel
; APPLICANT: Snyder, Benjamin
; APPLICANT: Reddy, Vermuri, B.
; APPLICANT: Wel, Chamer
; TITLE OF INVENTION: A cDNA Genomic Hybrid Sequence Encoding APP770
; Patent No. 5604131
; TITLE OF INVENTION: Containing a Genomic DNA Insert of the KI and OX-2 Regions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,702
; FILING DATE: 17-SEPT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: TSI121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-873-8794
; TELEFAX: (404)-873-8795
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-123-702-34

```

```
Query Match          76.08; Score 11.4; DB 1; Length 30;
Best Local Similarity 92.38; Pred. No. 4.9e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy	3	ccagtgagtct	15
Db	6	CCCAGGTAAGTCT	18

RESULT 11  
US-09-015-876-20  
; Sequence 20, Application US/09015876  
; Patent No. H001825  
; GENERAL INFORMATION:  
; APPLICANT: ROMANO, JOSEPH W  
; APPLICANT: SHURTLEIFF, ROXANNE  
; APPLICANT: WILLIAMS, KIMBERLY G  
; TITLE OF INVENTION: ISOTHERMAL TRANSCRIPTION BASED ASSAY  
; TITLE OF INVENTION: FOR THE DETECTION OF HTLV I AND HTLV II  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: AKZO NOBEL PATENT DEPT.  
STREET: 1300 PICCARD DRIVE, SUITE 206  
CITY: ROCKVILLE  
STATE: MARYLAND  
COUNTRY: US  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/015.876  
FILING DATE: 01-MAR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KLESNER, SHARON N  
REGISTRATION NUMBER: 36,335  
REFERENCE/DOCKET NUMBER: T/98353  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-948-7400  
TELEFAX: 301-948-9751  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-09-015-876-20

Query Match 73.3%; Score 11; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.1e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cccaggtgagt 13  
|||||  
Db 10 CCCAGGTGAGT 20

## RESULT 12

US-08-609-443B-53  
Sequence 53, Application US/08609443B  
Patent No. 5840693  
GENERAL INFORMATION:  
APPLICANT: ERIKSSON, Ulf  
APPLICANT: OLOFSSON, Birgitta  
APPLICANT: ALITALO, Kari  
APPLICANT: PAJUSOLA, Katri  
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND  
TITLE OF INVENTION: DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/609,443B  
FILING DATE: 01-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/397,651  
FILING DATE: 01-MAR-1995  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/469,427  
FILING DATE: 06-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/569,063  
FILING DATE: 06-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 1064/41979C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-609-443B-53

Query Match 73.3%; Score 11; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.1e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 caggtgagtct 15  
|||||  
Db 8 CAGGTGAGTCT 18

## RESULT 13

US-09-273-565-86/C  
Sequence 86, Application US/09273565A  
Patent No. 6166190  
GENERAL INFORMATION:  
APPLICANT: FUJIWARA, TSUTOMU  
APPLICANT: WATANABE, TAKESHI  
APPLICANT: HORIE, MASATO  
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME  
FILE REFERENCE: Q-53599  
CURRENT APPLICATION NUMBER: US/09/273,565A  
CURRENT FILING DATE: 1999-03-22  
EARLIER APPLICATION NUMBER: 09/055,699  
EARLIER FILING DATE: 1998-04-07  
EARLIER APPLICATION NUMBER: 08/820,170  
EARLIER FILING DATE: 1997-03-19  
EARLIER APPLICATION NUMBER: JP 63410/1996  
EARLIER FILING DATE: 1996-03-19  
EARLIER APPLICATION NUMBER: JP 69163/1997  
EARLIER FILING DATE: 1997-03-05  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 86  
LENGTH: 21  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: C1 Primer  
US-09-273-565-86

Query Match 73.3%; Score 11; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 caggtgagtct 15  
|||||  
Db 13 CAGGTGAGTCT 3

## RESULT 14

US-08-456-103-4  
; Sequence 4, Application US/08456103  
; Patent No. 5744304  
; GENERAL INFORMATION:  
; APPLICANT: Munford, Robert S.  
; TITLE OF INVENTION: INFLAMMATION-INDUCED EXPRESSION OF A  
; TITLE OF INVENTION: RECOMBINANT GENE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: United States of America  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/456,103  
; FILING DATE: Submitted Herewith  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: UTSD:406/KIT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
US-08-456-103-4

Query Match 73.3%; Score 11; DB 1; Length 28;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 cccaggtgagt 13  
Db 16 CCCAGGTGAGT 26

RESULT 15  
US-09-067-908-4  
; Sequence 4, Application US/09067908  
; Patent No. 5851822  
; GENERAL INFORMATION:  
; APPLICANT: Munford, Robert S.  
; TITLE OF INVENTION: INFLAMMATION-INDUCED EXPRESSION OF A  
; TITLE OF INVENTION: RECOMBINANT GENE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: United States of America  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/067,908  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/456,103  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: UTSD:406/KIT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
US-09-067-908-4

Query Match 73.3%; Score 11; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 cccaggtgagt 13  
Db 16 CCCAGGTGAGT 26

Search completed: October 2, 2001, 16:03:55  
Job time: 14599 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:48 ; Search time 876.95 Seconds

(without alignments)  
10.740 Million cell updates/sec

Title: US-09-757-100B-31

Perfect score: 15

Sequence: 1 taccagggtgagtct 15

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SID58/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	15	22	AAC65563 Human focal adhesi
2	15	100.0	20	22	AAC65543 Human focal adhesi
3	12.4	82.7	21	18	AAV58077 ICAM-1 antisense o
4	12.4	82.7	21	19	AAV38613 Human ICAM-1, E-se
5	12.4	82.7	22	16	AAQ82668 Chromosome 11 (loc
6	12.4	82.7	33	22	AAF98212 C neofomans strai
7	12	80.0	20	19	AAV28523 Blackcurrant rever
8	12	80.0	32	19	AAV28528 Human S1 receptor
9	11.8	78.7	34	19	AAV15506 Primer for adeno-a
10	11.8	78.7	48	20	AAV82665 Target oligonucleo
11	11.4	76.0	20	22	AAF62963 Mouse PEPCK-cytoso

c 12	11.4	76.0	25	21	AAA72481	Soybean chalcone r
c 13	11.4	76.0	46	11	AAO04939	Oligonucleotide ca
c 14	11	73.3	20	17	AAV37941	VEGF-B186 exon 5 b
c 15	11	73.3	20	17	AAV37943	VEGF-B167 exon 5 b
c 16	11	73.3	20	20	AAV63604	Human VEGF-B186 ex
c 17	11	73.3	20	22	AAV55942	Human eukaryotic i
c 18	11	73.3	21	18	AAV01904	Skeletal muscle sp
c 19	11	73.3	23	20	AAV08933	GATA-1 locus spllc
c 20	11	73.3	28	19	AAV22641	Primer used to amp
c 21	10.8	72.0	20	19	AAV57078	Human Notch3 mutan
c 22	10.8	72.0	21	19	AAV62658	Humanised antibody
c 23	10.8	72.0	21	19	AAV62659	Humanised antibody
c 24	10.8	72.0	21	19	AAV22891	Humanised LO-CD2a
c 25	10.8	72.0	21	19	AAV22892	Humanised LO-CD2a
c 26	10.8	72.0	21	20	AAV210191	PCR primer used to
c 27	10.8	72.0	21	20	AAV210192	PCR primer used to
c 28	10.8	72.0	23	21	AAV10299	HIV-1 LTR (long te
c 29	10.8	72.0	26	19	AAV56420	Human ICAM-R cDNA
c 30	10.8	72.0	26	20	AAV69192	Humanised ICR-1.1
c 31	10.8	72.0	26	21	AAV97162	PCR primer 8 used
c 32	10.8	72.0	26	21	AAV08308	Humanised ICR-1.1
c 33	10.8	72.0	26	21	AAV24334	Humanised murine a
c 34	10.8	72.0	28	17	AAV35702	3' VH primer VHP3
c 35	10.8	72.0	29	19	AAV00357	Bacillus thuringie
c 36	10.8	72.0	29	22	AAV73275	Oligonucleotide #6
c 37	10.8	72.0	30	18	AAV44073	Human vascular end
c 38	10.8	72.0	30	20	AAV81019	Humanised 340 Vh a
c 39	10.8	72.0	30	21	AAV15752	Human VEGF-3' PC
c 40	10.8	72.0	37	20	AAV23426	Human HAPO8 RACE P
c 41	10.8	72.0	37	21	AAV40044	PCR primer FOGIVHF
c 42	10.8	72.0	40	19	AAV56432	Human ICAM-R cDNA
c 43	10.8	72.0	40	19	AAV54868	PCR primer 93 used
c 44	10.8	72.0	40	20	AAV21887	Primer for antibod
c 45	10.8	72.0	40	20	AAV69200	Chimeric ICR-1.1 a

#### ALIGNMENTS

RESULT 1  
AAC65563  
ID AAC65563 standard; DNA; 15 BP.  
XX  
AC AAC65563;  
XX  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human focal adhesion kinase antisense sequence #29.  
XX  
KW Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphothioate; ss.  
XX  
OS Homo sapiens.  
XX  
PN US6133031-A.  
XX  
PD 17-OCT-2000.  
XX  
PF 19-AUG-1999; 99US-0377310.  
XX  
PR 19-AUG-1999; 99US-0377310.  
XX  
PA (ISIS-) ISIS PHARM INC.  
XX  
PI Monia BP, Gaarde WA;  
XX  
DR WPI; 2001-006141/01.  
XX  
PT New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -

PS Claim 15; Column 25; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
 CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
 CC is involved in integrin-mediated signal transduction, and is implicated  
 CC in cancer, particularly colon, breast and oral tumours, embryonic  
 CC development disorders, angiogenic disorders and wound healing. The  
 CC antisense sequences, including the one shown here, can be used in the  
 CC treatment of all of these.

XX Sequence 15 BP; 3 A; 4 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taccagggtgagctc 15

Db 1 taccagggtgagctc 15

RESULT 2

AAC65543  
 ID AAC65543 standard; DNA; 20 BP.

XX AAC65543;

DT 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #9.

XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
 KW embryonic development disorder; angiogenic disorder; wound healing;  
 KW antisense; phosphorothioate; ss.

XX Homo sapiens.

XX US6133031-A.

XX 17-OCT-2000.

PF 19-AUG-1999; 99US-0377310.

PR 19-AUG-1999; 99US-0377310.

XX (ISIS-) ISIS PHARM INC.

PI Monia BP, Gaarde WA;

XX WPI; 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase  
 PT expression, especially useful for inhibiting retinal  
 PT neovascularization, or for diagnosing and treating e.g. colon cancer -

PS Claim 15; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
 CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
 CC is involved in integrin-mediated signal transduction, and is implicated  
 CC in cancer, particularly colon, breast and oral tumours, embryonic  
 CC development disorders, angiogenic disorders and wound healing. The  
 CC antisense sequences, including the one shown here, can be used in the  
 CC treatment of all of these.

XX Sequence 20 BP; 5 A; 4 C; 6 G; 5 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 15; DB 22; Length 20;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taccagggtgagctc 15

Db 3 taccagggtgagctc 17

RESULT 3

AAT58077/c

ID AAT58077 standard; DNA; 21 BP.

XX AAT58077;

XX 18-MAR-1997 (first entry)

XX ICAM-1 antisense oligonucleotide #7.

XX Antisense; pre-mRNA; mature mRNA; vascular defect; tissue defect;  
 KW human intercellular adhesion molecule-1; ICAM-1; inflammation;  
 KW adult respiratory distress syndrome; multiple organ failure; GM1594;  
 KW septic shock; ss.

XX Synthetic.

XX US5580969-A.

XX 03-DEC-1996.

XX 24-JUL-1992; 92US-0918259.

XX 12-OCT-1993; 93US-0136118.

XX 24-JUL-1992; 92US-0918259.

XX (USNA) US SEC OF NAVY.

XX Bradley MO, Hoke GD, Lee C, Williams TJ;

XX WPI; 1997-033603/03.

XX Anti-sense oligo:nucleotide(s) for blocking ICAM-1 mRNA translation  
 PT - for treating septic shock, adult respiratory distress syndrome  
 PT etc.

XX Claim 1; Column 21; 16pp; English.

XX The sequences given in AAT58071-85 represent oligonucleotides which are  
 CC antisense to sequences contained in the pre-mRNA or mature mRNA  
 CC transcript of human intercellular adhesion molecule-1 (ICAM-1).  
 CC These oligonucleotides may be used for treating septic shock and the  
 CC manifestations of septic shock, e.g. inflammation, and vascular and  
 CC tissue defects. They are also useful in the treatment of septic  
 CC shock associated diseases, e.g. adult respiratory distress syndrome,  
 CC multiple organ failure etc.

XX Sequence 21 BP; 5 A; 9 C; 5 G; 2 T; 0 other;

Query Match

Best Local Similarity 82.7%; Score 12.4; DB 18; Length 21;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 taccagggtgagtc 14

Db 18 TTCCAGGTGAGTC 5

RESULT 4

AAV38613/c

ID AAV38613 standard; DNA; 21 BP.

XX AAV38613;

XX 13-OCT-1998 (first entry)

XX Human ICAM-1, E-selectin, VCAM-1 antisense oligonucleotide.



```

XX ICAM-1; intracellular adhesion molecule-; E-selectin; VCAM-1;
KW vascular cell adhesion molecule-1; antiseptic; inflammatory;
KW disease; treatment; septic shock; psoriasis; wounds; burns; acne;
KW arthritis; organ rejection; inhibition; expression; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
XX WO9824797-A1.
XX
XX 11-JUN-1998.
XX
XX 02-DEC-1996; 96WO-US19194.
XX
XX 02-DEC-1996; 96WO-US19194.
XX
XX (DYAD-) DYAD PHARM CORP.
XX
XX Bradley MO, Hoke GD, Lee C, Williams TJ;
XX WPI; 1998-333253/29.
XX
XX Antisense oligonucleotides to ICAM-1, E-selectin or VCAM-1 - useful
PT for treating diseases having an inflammatory component, e.g.
PT psoriasis, wounds and septic shock
XX
XX Claim 8; Page 40; 48pp; English.
XX
XX The sequence is that of an antisense oligonucleotide which is
CC substantially complementary to at least a portion of the pre-
CC or mature RNA transcript of human intracellular adhesion molecule
CC (ICAM), E-selectin or vascular cell adhesion molecule (VCAM).
CC It can be used to inhibit expression of these proteins. Inhibition
CC of these proteins forms the basis for treatment of conditions and
CC diseases that have an inflammatory component, e.g. acne, psoriasis,
CC arthritis, organ rejection, wounds, burns, septic shock or
CC inflammatory complications of septic shock.
XX
XX Sequence 21 BP; 5 A; 9 C; 5 G; 2 T; 0 other;
SQ

```

Query Match 82.7%; Score 12.4; DB 19; Length 21;  
Best Local Similarity 92.9%; Pred. No. 4.7e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 taccacagtgagtc 14
    | |||||
Db 18 TTCCACGGTGAGTC 5

```

RESULT 5  
AAQ82668/C  
ID AAQ82668 standard; DNA; 22 BP.  
AC  
AC AAQ82668;  
XX  
XX 14-SEP-1995 (first entry)  
XX  
XX Chromosome 11 (locus LDHC) STS primer LDHC-2.  
DE  
XX  
XX sequence sampled mapping; genomic analysis; complex genome mapping;  
KW cosmid library; chromosome 11; sequence tagged site; STS analysis; ss.  
XX  
XX Synthetic.  
XX  
XX WO9429486-A.  
XX  
XX 22-DEC-1994.  
XX  
XX 15-JUN-1994; 94WO-US06810.  
XX  
XX 15-JUN-1993; 93US-0078471.  
XX  
XX

```

PR 07-SEP-1993; 93US-0117952.
XX
XX (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
XX Evans GA, Smith MW;
XX
XX WPI; 1995-036508/05.
XX
XX Sequencing complex genomes, present as fragments in a cosmid
PT library - by sequencing end-specific nucleotides of each clone
PT then correlating with spatial relationship of cosmid, esp. for
XX mammalian chromosomes.
XX
XX Example 4; Page 92; 128pp; English.
XX
XX Sequences were determined from the ends of chromosome 11-specific
CC cosmid by automated sequencing without intermediate subcloning.
CC A sample of 371 DNA sequence fragments were determined and of
CC these, 277 were suitable for STS primer prediction by computer
CC analysis (using the "Primer" program available from E.Lander, MIT).
CC The STSs and cosmids were mapped by in situ hybridisation, somatic
CC cell hybrid analysis or both. Using this method, 370 STSs specific
CC for human chromosome 11 were generated and most of them were
CC regionally mapped. This procedure illustrates a novel method for
CC sequencing complex genomes, designated "sequence sampled mapping".
CC The sequence sampled mapping method is useful for the completion of
CC high density sequence-based maps, and ultimately, for the complete
CC sequencing of genomic DNA directly from cosmid clones.
XX See AAQ82001-Q82706 and AAQ91325-Q91358 for STS primers.
XX
XX Sequence 22 BP; 8 A; 5 C; 4 G; 5 T; 0 other;
SQ

```

Query Match 82.7%; Score 12.4; DB 16; Length 22;  
Best Local Similarity 92.9%; Pred. No. 4.7e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 2 accacagtgagtc 15
    | |||||
Db 14 ATCCAGGTGAGTCT 1

```

RESULT 6  
AAF98212  
ID AAF98212 standard; DNA; 33 BP.  
XX  
XX AAF98212;  
XX  
XX 05-JUN-2001 (first entry)  
XX  
XX C neoformans strain discrimination probe #30.  
DE  
XX  
XX Pathogen; yeast; intergenic spacer region; IGS1; PCR primer; probe; ss.  
XX  
XX Cryptococcus neoformans.  
XX  
XX WO200123616-A2.  
XX  
XX 05-APR-2001.  
XX  
XX 29-SEP-2000; 2000WO-US26758.  
XX  
XX 29-SEP-1999; 99US-0156598.  
XX  
XX (GENE-) GENETIC VECTORS INC.  
XX (FELL/) FELL J.  
XX (DIAZ/) DIAZ M.  
XX  
XX Fell J, Diaz M, McCabe M;  
XX  
XX WPI; 2001-258138/26.  
XX  
XX Novel assemblage useful for discriminating among pathogenic yeasts,  
XX

PT comprises two universal primers adapted for nucleic acid amplification  
 protocol

PS Claim 6; Page 18; 88pp; English.

XX The present invention describes an assemblage comprising two primers,  
 CC each of which can be used to amplify the intergenic spacer region IGS1  
 CC from one of various strains of the yeast *Cryptococcus neoformans*. A  
 CC number of primers and probes are provided, as are the sequences of the  
 CC IGS1 for 91 *C. neoformans* strains. This is useful in the discrimination  
 CC of pathogenic yeasts, and the sequences can be used to construct a  
 CC database having the same purpose. The present sequence is a probe or  
 CC primer described in the invention.

XX Sequence 33 BP; 7 A; 6 C; 13 G; 7 T; 0 other;  
 SQ

Query Match 82.7%; Score 12.4; DB 22; Length 33;

Best Local Similarity 92.9%; Pred. No. 4.8e+02;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 acccaggtagtct 15

Db 17 acccaggtagtat 30

RESULT 7

AAV28523

ID AAV28523 standard; DNA; 20 BP.

XX AAV28523;

DT 28-AUG-1998 (first entry)

XX Blackcurrant reversion virus RNA2 3' proximal fragment primer 5.

DE Blackcurrant reversion disease; BRV; rna2; diagnosis; Ribes; PCR;

XX primer; ss.

OS Synthetic.

XX Blackcurrant reversion virus.

XX WO9810100-A1.

XX 12-MAR-1998.

PF 01-SEP-1997; 97WO-FI00507.

XX 05-SEP-1996; 96FI-0003474.

XX (ABOA-) ABOATECH OY AB.

XX Latvala S, Lehto K, Lemmetty A, Susi P;

XX WPI; 1998-193642/17.

XX Diagnosing blackcurrant reversion disease in plants e.g.

PT blackcurrant - using reverse transcriptase-PCR with primers

PT amplifying cDNA fragment complementary to fragment of new

PT blackcurrant reversion virus

XX Example 2; Page 27; 38pp; English.

XX Primer 5 and primer 6 (see AAV28524) are designed to amplify a 481

CC nucleotide DNA fragment initiating 265 nucleotides upstream of the

CC polyA tail of RNA2 of a blackcurrant reversion virus (BRV) Finnish

CC isolate genome (see AAV28525). They were used with primer 1,2 (see

CC AAV28521-22), designed to amplify a 210 nucleotide fragment, to

CC detect different isolates of BRV originating from widely different

CC geographic locations. Primer pair 1,2 amplified the expected

CC virus-specific fragments from all tested virus isolates, and the

CC primer pair 5,6 amplified the expected virus-specific fragment from

CC nearly all the tested virus isolates, indicating that the viral

CC sequences detecting by these primer pairs are well conserved in all  
 CC isolates, including the common (E) and strong (R) forms of the  
 CC reversion disease. The invention provides methods and kits  
 CC for the diagnosis of blackcurrant reversion disease using RT-PCR.

XX Sequence 20 BP; 7 A; 5 C; 6 G; 2 T; 0 other;

Query Match

Best Local Similarity 80.0%; Score 12; DB 19; Length 20;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 acccaggtagt 13

Db 8 acccaggtagt 19

RESULT 8

AAAT99528

ID AAT99528 standard; DNA; 32 BP.

XX AAT99528;

DT 21-MAY-1998 (first entry)

XX Human ST receptor PCR primer.

DE ST receptor; heat stable toxin receptor; colorectal cancer; tumour;

XX metastasis; diagnosis; human; PCR; primer; ss.

OS Synthetic.

XX Homo sapiens.

XX WO9742506-A1.

XX 13-NOV-1997.

PF 02-MAY-1997; 97WO-US07467.

XX 03-MAY-1996; 96US-0016564.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Carithers SL, Waldman SA;

XX WPI; 1998-008454/01.

XX Determining whether an individual has metastasised colorectal cancer

PT cells and origin of tumour cells - by detecting presence of

PT heat-stable toxin receptor on cells in a sample

XX Claim 14; Page 54; 62pp; English.

XX Claimed PCR primers (see AAT99462-T99531) hybridise to sequences that

CC encode the extracellular domain of human heat-stable toxin (ST)

CC receptor protein (see AAW37371), a highly specific marker for

CC metastasised colorectal cancer cells. PCR using these primers

CC provides specific and sensitive detection of human ST receptor

CC expression. A specific primer pair comprises the primers given in

CC AAT99528 and AAT99529. Claimed in vitro methods for determining whether

CC or not (i) an individual has metastasised colorectal cancer cells,

CC or (ii) a tumour cell is a colorectal cancer cell comprise the

CC steps of examining a sample of extraintestinal tissue and/or body

CC fluids or tumour cells from an individual to determine whether ST

CC receptor protein is being expressed by cells in the sample.

CC Expression is determined by immunoassay or by PCR using primers

CC that selectively amplify ST receptor cDNA or mRNA (see also

CC AAT97229).

XX Sequence 32 BP; 6 A; 11 C; 9 G; 6 T; 0 other;

Query Match

80.0%; Score 12; DB 19; Length 32;

```

Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cccaggtgagtc 14
DB 12 cccaggtgagtc 23

RESULT 9
AAV15506
ID AAV15506 standard; DNA; 34 BP.
XX
AC AAV15506;
XX
DT 11-JUN-1998 (first entry)
XX
DE
XX
DE
XX
KW AAV; vector; rep 68/78; PCR primer; gene therapy; ss.
XX
OS Synthetic.
OS Adeno associated virus.
XX
PN WO9749824-A1.
XX
PD 31-DEC-1997.
XX
PF 24-JUN-1997; 97WO-DE01333.
XX
PR 24-JUN-1996; 96DE-1025188.
XX
PA (MEDI-) MEDIGENE GMBH.
PA (MEDI-) MEDIGENE AG.
XX
PI Bogedain C, Hallek M, Maass G;
XX
DR WPI; 1998-086641/08.
XX
PT System of adeno-associated viral vector and rep 68/78 sequence of
PT this virus - expression of which is delayed until replication of
PT viral DNA has started, provides large scale production of vectors
PT for gene therapy
XX
PS Example; Page 7; 20pp; German.
XX
CC The present sequence was used in the development of a novel system,
CC comprising an adeno-associated virus (AAV) vector containing a
CC foreign DNA, and the rep 68/78 sequence of AAV, the expression of
CC which is delayed. The components may be present in cis (in a single
CC agent) or in trans (in separate agents).
CC The system is used for production of AAV vectors, particularly for
CC gene therapy. The foreign DNA may encode a therapeutic protein,
CC e.g. interferon, interleukin, growth factor, coagulation factor or
CC metabolic enzyme, particularly one that increases the
CC immunogenicity of tumour cells, and/or a diagnostic protein. The
CC system produces AAV vectors on a large scale. It is based on the
CC discovery that the rep68 and 78 proteins interfere with replication
CC of AAV DNA, and that this interference is overcome by delaying
CC expression of these proteins.
XX
SQ Sequence 34 BP; 4 A; 13 C; 9 G; 8 T; 0 other;

Query Match 78.7%; Score 11.8; DB 19; Length 34;
Best Local Similarity 86.7%; Pred. No. 1.1e+03;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taccaggtgagtc 15
DB 6 taccaggtgagtc 20

RESULT 10
AAV82665/c
ID AAV82665 standard; DNA; 48 BP.
XX
AC AAV82665;
XX
DT 25-FEB-1999 (first entry)
XX
DE
XX
DE
XX
KW Target oligonucleotide JN5.
KW Detector oligonucleotide; hairpin structure; fluorescence;
KW primer extension; hybridization; signal primer; frameshift mutation;
KW ss.
XX
OS Synthetic.
XX
PN EP881302-A2.
XX
PD 02-DEC-1998.
XX
PF 28-MAY-1998; 98EP-0109682.
XX
PR 30-MAY-1997; 97US-0865675.
XX
PA (BECT ) BECTON DICKINSON & CO.
XX
PI Linn PC, Nadeau JG, Pitner BJ, Schram JL;
XX
DR WPI; 1999-001406/01.
XX
PT New detector oligo:nucleotide having base-paired region carrying
PT quenched dyes - where dyes become fluorescent when region is
PT unpaired, useful for detection and amplification of target nucleic
PT acid
XX
PS Example 2; Page 12; 20pp; English.
XX
CC Target oligonucleotides AAV82665-68 were designed to hybridise to the
CC detector oligonucleotide of the invention. The detector oligonucleotide
CC comprises a single-stranded target-binding region and an
CC intramolecularly base-paired secondary structure linked to two dyes
CC (donor and acceptor fluorophores). In the secondary structure,
CC fluorescence of the donor is quenched, but when it is linearised or
CC unfolded a change in some fluorescence parameter becomes detectable. A
CC target nucleic acid is detected by hybridizing it to a detector
CC oligonucleotide in which the secondary structure is 5' to the target
CC binding region, primer extension to produce a complementary strand using
CC the secondary structure as template, resulting in linearization or
CC unfolding of it and detecting a change in fluorescence. The detector
CC oligonucleotide are used to detect (by primer extension and
CC hybridization) and amplify (as signal primer) target sequences, e.g. for
CC detecting frameshift mutations.
XX
SQ Sequence 48 BP; 15 A; 8 C; 10 G; 15 T; 0 other;

Query Match 78.7%; Score 11.8; DB 20; Length 48;
Best Local Similarity 86.7%; Pred. No. 1.1e+03;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 taccaggtgagtc 15
DB 41 TACTCAGATGAGTCT 27

RESULT 11
AAV62963/c
ID AAV62963 standard; DNA; 20 BP.
XX
AC AAV62963;
XX
DT 08-MAY-2001 (first entry)
XX
DE Mouse PEPCCK-cytosolic antisense oligonucleotide ISIS 113360.

```

XX Mouse; antiinflammatory; cytostatic; antisense gene therapy;  
 KW phosphoenol pyruvate carboxykinase-cytosolic; PEPCK-cytosolic;  
 XX infection; inflammation; tumour formation; phosphorothioate; ss.  
 XX  
 OS Mus musculus.  
 XX  
 PN US6187545-B1.  
 XX  
 PD 13-FEB-2001.  
 XX  
 PF 21-JAN-2000; 2000US-0488671.  
 XX  
 PR 21-JAN-2000; 2000US-0488671.  
 XX  
 PA (ISIS-) ISIS PHARM INC.  
 XX  
 PI McKay R, Butler MM, Wyatt J, Cowsert LM;  
 XX  
 DR WPI; 2001-190979/19.  
 XX  
 DR Antisense compound capable of modulating the expression of phosphoenol  
 XX pyruvate carboxykinase-cytosolic, useful for preventing or delaying  
 PT infection, inflammation or tumor formation -  
 PT  
 XX Example 17; Column 44; 64pp; English.  
 PS  
 XX The present sequence is one of a number of antisense compounds of up to  
 CC 30 nucleobases in length that are capable of inhibiting the expression of  
 CC phosphoenol pyruvate carboxykinase-cytosolic (PEPCK-cytosolic). The  
 CC antisense compounds are useful for inhibiting the expression of  
 CC PEPCK-cytosolic in cells or tissues. They are commonly used as research  
 CC reagents and in diagnostics, e.g. to elucidate the function of particular  
 CC genes. They are also useful for distinguishing between functions of  
 CC various members of a biological pathway and for research use. The  
 CC antisense compounds are also useful prophylactically, e.g. to prevent or  
 CC delay infection, inflammation or tumour formation. The present sequence  
 CC is a chimeric phosphorothioate oligonucleotide with 2'-MOE wings and a  
 CC deoxy gap.  
 XX  
 SQ Sequence 20 BP; 5 A; 6 C; 5 G; 4 T; 0 other;

Query Match 76.0%; Score 11.4; DB 22; Length 20;  
 Best Local Similarity 92.3%; Pred. No. 1.7e+03;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ccaggtgagtct 15  
 || |||||  
 DB 15 CCAAGGTGAGTCT 3

RESULT 12  
 AAA72481/c  
 ID AAA72481 standard; DNA; 25 BP.  
 XX  
 AC AAA72481;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Soybean chalcone reductase PCR primer, SEQ ID NO:65.  
 XX  
 KW Soybean; chalcone reductase; phenylpropanoid pathway;  
 KW isoflavone synthase; isoflavonoid biosynthesis; defence response;  
 KW attractant; repellent; signal compound; antibiotic; transgenic plant;  
 KW transgenic seed; PCR primer; ss.  
 XX  
 OS Glycine max.  
 XX  
 PN WO200044909-A1.  
 XX  
 PD 03-AUG-2000.  
 XX

PF 26-JAN-2000; 2000WO-US01772.  
 XX  
 PR 27-JAN-1999; 99US-0117769.  
 PR 20-JUL-1999; 99US-0144783.  
 PR 24-SEP-1999; 99US-0156094.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Fader GM, Jung W, McGonigle B, Odell JT, Yu X;  
 XX  
 PR WPI; 2000-543395/49.  
 DR  
 XX  
 PT Nucleic acids encoding isoflavonoid synthases, useful for producing  
 PT transgenic plants with increased production of isoflavonoids which are  
 PT involved in defense against phytopathogenic microorganisms -  
 XX  
 PS Example 14; Page 52; 157pp; English.  
 XX  
 CC Sequences AAA72437, AAA72449-A72461, AAA72468 and AAA72474-A72477  
 CC represent cDNAs encoding novel plant isoflavone synthases (AAB21052 and  
 CC AAB21054- AAB21071). Also disclosed is the soybean cytochrome P450  
 CC monooxygenase CYP93C1, encoded by a known sequence (AAA72444, NCBI No.  
 CC 2739005), which was identified in the present invention as having  
 CC isoflavone synthase activity. The invention also relates to expression  
 CC constructs, transformed host cells, and transgenic plants and seeds  
 CC comprising the novel cDNA sequences of the invention. The invention also  
 CC encompasses methods of altering isoflavone synthase expression in a host  
 CC cell, altering isoflavonoid levels in a plant, and identifying nucleic  
 CC acids encoding other plant isoflavone synthases. Isoflavone synthase  
 CC plays a key role in the biosynthesis of isoflavonoids. Isoflavonoids are  
 CC a class of secondary metabolites mainly produced in leguminous plants by  
 CC a branch of the phenylpropanoid pathway. Isoflavone synthase catalyses  
 CC the first step in the branch of this pathway that commits metabolic  
 CC intermediates to the synthesis of isoflavonoids. Isoflavonoids  
 CC participate in the defence response of legumes against phytopathogenic  
 CC microorganisms and are also involved in symbiotic relationships between  
 CC the roots of legumes and rhizobial bacteria which eventually result in  
 CC nodulation and nitrogen-fixation. They have also been shown to act as  
 CC antibiotics, repellents, attractants, and signal compounds, and  
 CC consumption of legume isoflavonoid is associated with health benefits  
 CC in humans. The novel isoflavonoid synthases, and nucleic acids encoding  
 CC them, are useful for altering the levels of isoflavonoids produced in  
 CC legumes such as soybean, and for the production of isoflavonoids in  
 CC plants which do not naturally produce them (e.g., maize, rice, wheat).  
 CC Sequences AAA72478-A72481 represent PCR primers used in an  
 CC exemplification of the invention to amplify DNA encoding soybean chalcone  
 CC reductase (an enzyme that acts upstream of isoflavone synthase in the  
 CC phenylpropanoid pathway).  
 XX  
 SQ Sequence 25 BP; 6 A; 7 C; 5 G; 7 T; 0 other;

Query Match 76.0%; Score 11.4; DB 21; Length 25;  
 Best Local Similarity 92.3%; Pred. No. 1.7e+03;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ccaggtgagtct 15  
 |||||  
 DB 17 CCCAGGTGAGTTT 5

RESULT 13  
 AAQ04939/c  
 ID AAQ04939 standard; DNA; 46 BP.  
 XX  
 AC AAQ04939;  
 XX  
 DT 24-OCT-1990 (first entry)  
 XX  
 DE Oligonucleotide carrying mutation for factor VIII gene.  
 XX  
 KW Human factor VIII analogue; ss.  
 XX



PR 06-JUN-1995; 95US-0459427.  
XX  
XX (LUDW-) LUDWIG INST CANCER RES.  
PA (UYHE-) UNIV HELSINKI LICENSING LTD OY.  
XX  
XX  
XX Alitalo K, Eriksson U, Olofsson B, Pajusola K;  
XX WPI; 1996-412582/41.  
XX  
XX Vascular endothelial growth factor VEGF-B proteins - useful to  
PT accelerate angiogenesis in wound healing, also related nucleic acid  
PT and antibodies for cancer diagnosis  
XX  
PS Example 7; Page 28; 107pp; English.  
XX  
XX AAT37933-T37946 represent the intron/exon boundaries for the human  
CC vascular endothelial growth factor (VEGF) proteins of the invention (see  
CC AAW04829, and AAW04831), which promote endothelial or mesodermal cell  
CC proliferation. VEGF is also a glycosylated cationic dimer, and is  
CC sometimes referred to as vascular permeability factor (VPF). VEGF has  
CC diverse effects, depending on the specific biological context in which  
CC it is found. VEGF is a potent endothelial cell mitogen, and directly  
CC contributes to induction of angiogenesis in vivo by promoting  
CC endothelial cell growth during normal embryonic development, wound  
CC healing, and tissue regeneration/reorganisation. The VEGF proteins of  
CC the invention share the angiogenic and other properties of VEGF, but are  
CC distributed and expressed in tissues differently to VEGF. The proteins  
CC can therefore be used to accelerate angiogenesis in wound healing.  
CC Antibodies against the proteins can be used for inhibiting angiogenesis.  
CC The antibodies can also be used diagnostically to quantitatively detect  
CC VEGF-B. Primers complementary to the coding sequences for the proteins  
CC of the invention can also be used to detect VEGF-B coding sequences.  
CC Quantification of VEGF-B in cancer biopsy specimens may be useful as an  
CC indicator of metastatic risk. VEGF-B expression in a cell can be  
CC retarded using antisense sequences direct against the VEGF coding  
CC sequences, this is especially useful in retarding VEGF expression in  
CC tumour cells.  
XX  
SQ Sequence 20 BP; 5 A; 4 C; 6 G; 5 T; 0 other;

Query Match 73.3%; Score 11; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 caggtgagctct 15  
|||||  
Db 8 caggtgagctct 18

Search completed: October 2, 2001, 16:18:48  
Job time: 15492 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:03:54 ; Search time 417.38 Seconds  
(without alignments)  
6.804 Million cell updates/sec

Title: US-09-757-100B-30  
Perfect score: 15  
Sequence: 1 taagcagctgccatt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 460742

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	3	US-09-377-310-30
2	15	100.0	20	3	US-09-377-310-10
c 3	13.4	89.3	18	1	US-08-378-761A-45
c 4	13.4	89.3	18	1	US-08-485-286-45
c 5	12	80.0	29	3	US-08-826-964-4
6	11.8	78.7	50	1	US-08-171-389-486
7	11.8	78.7	50	1	US-08-171-389-487
8	11.8	78.7	50	1	US-08-123-936-486
9	11.8	78.7	50	1	US-08-123-936-487
10	11.8	78.7	50	2	US-08-475-228A-486
11	11.8	78.7	50	2	US-08-475-228A-487
12	11.8	78.7	50	3	US-08-482-080A-486
13	11.8	78.7	50	3	US-08-482-080A-487
14	11.8	78.7	50	5	PCT-US93-12388-486
15	11.8	78.7	50	5	PCT-US93-12388-487
16	11.4	76.0	20	2	US-08-832-658A-4
17	11.4	76.0	20	4	US-09-306-876A-5
18	11.4	76.0	20	4	US-09-306-876A-6
19	11.4	76.0	25	1	US-08-482-115B-26
20	11.4	76.0	25	2	US-08-472-802C-27
21	11.4	76.0	30	1	US-08-833-377-5
22	11.4	76.0	39	2	US-07-759-568-5
c 23	11.4	76.0	39	2	US-08-484-397A-24
c 24	11.4	76.0	39	2	US-08-834-655-15
c 25	11.4	76.0	39	3	US-09-363-574-15
c 26	11	73.3	22	2	US-08-479-614-21
c 27	11	73.3	28	1	US-08-479-852-12

c 28	11	73.3	28	1	US-08-479-852-64	Sequence 64, Appl
c 29	11	73.3	28	1	US-08-479-852-78	Sequence 78, Appl
c 30	11	73.3	28	1	US-08-479-852-92	Sequence 92, Appl
c 31	11	73.3	28	2	US-08-482-846-12	Sequence 12, Appl
c 32	11	73.3	28	2	US-08-482-846-64	Sequence 64, Appl
c 33	11	73.3	28	2	US-08-462-646-92	Sequence 78, Appl
c 34	11	73.3	28	2	US-08-462-646-92	Sequence 92, Appl
c 35	11	73.3	28	4	US-09-013-406-12	Sequence 12, Appl
c 36	11	73.3	28	4	US-09-013-406-64	Sequence 64, Appl
c 37	11	73.3	28	4	US-09-013-406-78	Sequence 78, Appl
c 38	11	73.3	28	4	US-09-013-406-92	Sequence 92, Appl
c 39	11	73.3	29	1	US-08-233-030-37	Sequence 37, Appl
c 40	11	73.3	35	1	US-08-343-682-12	Sequence 12, Appl
c 41	11	73.3	37	4	US-09-198-955A-21	Sequence 21, Appl
c 42	11	73.3	39	2	US-08-951-822-13	Sequence 13, Appl
c 43	11	73.3	39	4	US-09-173-043-28	Sequence 28, Appl
c 44	11	73.3	41	2	US-08-484-397A-19	Sequence 19, Appl
c 45	11	73.3	45	4	US-09-312-266-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1  
US-09-377-310-30  
; Sequence 30, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-30

Query Match 100.0%; Score 15; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 9.6; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 0;  
Qy 1 taagcagctgccatt 15  
Db 1 taagcagctgccatt 15

RESULT 2  
US-09-377-310-10  
; Sequence 10, Application US/09377310B  
; Patent No. 6133031  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Gaarde, William A.  
; TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: ISPH-0389  
; CURRENT APPLICATION NUMBER: US/09/377,310B  
; CURRENT FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-10

Query Match 100.0%; Score 15; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
|||||  
Db 3 taagcagctgccatt 17

RESULT 3  
US-08-378-761A-45/C  
; Sequence 45, Application US/08378761A  
; Patent No. 5635384  
; GENERAL INFORMATION:  
; APPLICANT: WALSH, TERENCE A  
; APPLICANT: HEY, TIMOTHY D  
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE  
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF  
; NUMBER OF SEQUENCES: 81  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ANDREA T. BORUCKI  
; STREET: 9330 ZIONSVILLE ROAD  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: US  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/378,761A  
; FILING DATE: 26-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BORUCKI, ANDREA T  
; REGISTRATION NUMBER: 33651  
; REFERENCE/DOCKET NUMBER: 38272B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317) 337-4846  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-378-761A-45

Query Match 89.3%; Score 13.4; DB 1; Length 18;  
Best Local Similarity 93.3%; Pred. No. 72;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
|||||  
Db 17 TAAGCAGCTGCAATT 3

RESULT 4  
US-08-485-286-45/c  
; Sequence 45, Application US/08485286  
; Patent No. 5646026  
; Patent No. 5646026 5646119  
; GENERAL INFORMATION:  
; APPLICANT: WALSH, TERENCE A

; APPLICANT: HEY, TIMOTHY D  
; APPLICANT: MORGAN, ALICE ER  
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE  
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF  
; TITLE OF INVENTION: USING  
; NUMBER OF SEQUENCES: 81  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ANDREA T. BORUCKI  
; STREET: 9330 ZIONSVILLE ROAD  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: US  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,286  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/378761  
; FILING DATE: 26-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BORUCKI, ANDREA T  
; REGISTRATION NUMBER: 33651  
; REFERENCE/DOCKET NUMBER: 38272B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317) 337-4846  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-485-286-45

Query Match 89.3%; Score 13.4; DB 1; Length 18;  
Best Local Similarity 93.3%; Pred. No. 72;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
|||||  
Db 17 TAAGCAGCTGCAATT 3

RESULT 5  
US-08-826-964-4/c  
; Sequence 4, Application US/08826964  
; Patent No. 6033663  
; GENERAL INFORMATION:  
; APPLICANT: Ketcham, Catherine M.  
; TITLE OF INVENTION: Nucleic Acids Encoding GDP-Fucose  
; TITLE OF INVENTION: Pyrophosphorylase  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,964  
; FILING DATE: 09-APR-1997



CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/015,241  
FILING DATE: 10-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 014137-009910US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-826-964-4

Query Match 80.08; Score 12; DB 3; Length 29;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 agcagctgccaat 14  
|||||  
Db 27 AGCAGCTGCCAT 16

RESULT 6  
US-08-171-389-486  
Sequence 486, Application US/08171389  
Patent No. 5578444  
GENERAL INFORMATION:  
APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.  
APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
APPLICANT: Fry, Kirk E.  
TITLE OF INVENTION: Sequence-Directed DNA Binding  
TITLE OF INVENTION: Molecules, Compositions and Methods  
NUMBER OF SEQUENCES: 641  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/171,389  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.

REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0175/G19P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 486:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Human T-cell Lymphotropic virus type  
INDIVIDUAL ISOLATE: III (HIV-1)  
US-08-171-389-486

Query Match 78.78; Score 11.8; DB 1; Length 50;  
Best Local Similarity 86.7%; Pred. No. 5.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 taagcagctgccatt 15  
|||||  
Db 24 TAAGCAGCTGCTTTT 38

RESULT 7  
US-08-171-389-487  
Sequence 487, Application US/08171389  
Patent No. 5578444  
GENERAL INFORMATION:  
APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.  
APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
APPLICANT: Fry, Kirk E.  
TITLE OF INVENTION: Sequence-Directed DNA Binding  
TITLE OF INVENTION: Molecules, Compositions and Methods  
NUMBER OF SEQUENCES: 641  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/171,389  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0175/G19P3

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 487:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Aids-associated retrovirus  
INDIVIDUAL ISOLATE: (arv-2;provincial)  
US-08-171-389-487

Query Match 78.7%; Score 11.8; DB 1; Length 50;  
Best Local Similarity 86.7%; Pred. No. 5.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
Db 24 TAAGCAGCTGCTTTT 38

RESULT 8  
US-08-123-936-486  
; Sequence 486, Application US/08123936  
; Patent No. 5726014  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; TITLE OF INVENTION: Screening Assay for the Detection of  
; TITLE OF INVENTION: DNA-Binding Molecules  
; NUMBER OF SEQUENCES: 640  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/123,936  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 486:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: RNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Human T-cell Lymphotropic virus type  
INDIVIDUAL ISOLATE: III (HIV-1)  
US-08-123-936-486

Query Match 78.7%; Score 11.8; DB 1; Length 50;  
Best Local Similarity 86.7%; Pred. No. 5.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
Db 24 TAAGCAGCTGCTTTT 38

RESULT 9  
US-08-123-936-487  
; Sequence 487, Application US/08123936  
; Patent No. 5726014  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; TITLE OF INVENTION: Screening Assay for the Detection of  
; TITLE OF INVENTION: DNA-Binding Molecules  
; NUMBER OF SEQUENCES: 640  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/123,936  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 487:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Aids-associated retrovirus  
INDIVIDUAL ISOLATE: (arv-2;provincial)  
US-08-123-936-487

Query Match 78.7%; Score 11.8; DB 1; Length 50;

Best Local Similarity 86.7%; Pred. No. 5.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
|||||  
Db 24 TAAGCAGCTGCTTTT 38

## RESULT 10

US-08-475-228A-486  
; Sequence 486, Application US/08475228A  
; Patent No. 5869241  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,228A  
FILING DATE: 06-JUN-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Stratford, Carol A.

REGISTRATION NUMBER: 34,444

REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P3D2

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 486:

SEQUENCE CHARACTERISTICS:

LENGTH: 50 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: RNA (genomic)

HYPOTHETICAL: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: Human T-cell Lymphotropic virus type

INDIVIDUAL ISOLATE: III (HIV-1)

US-08-475-228A-486

Query Match 78.7%; Score 11.8; DB 2; Length 50;  
Best Local Similarity 86.7%; Pred. No. 5.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
|||||  
Db 24 TAAGCAGCTGCTTTT 38

Qy 1 taagcagctgccatt 15  
|||||  
Db 24 TAAGCAGCTGCTTTT 38

## RESULT 11

US-08-475-228A-487  
; Sequence 487, Application US/08475228A  
; Patent No. 5869241  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,228A  
FILING DATE: 06-JUN-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Stratford, Carol A.

REGISTRATION NUMBER: 34,444

REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P3D2

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 487:

SEQUENCE CHARACTERISTICS:

LENGTH: 50 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: RNA (genomic)

HYPOTHETICAL: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: Aids-associated retrovirus

INDIVIDUAL ISOLATE: (arv-2;proviral)

US-08-475-228A-487

Query Match 78.7%; Score 11.8; DB 2; Length 50;  
Best Local Similarity 86.7%; Pred. No. 5.8e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
|||||  
Db 24 TAAGCAGCTGCTTTT 38



```
RESULT 14
PCT-US93-12388-486
; Sequence 486, Application PC/TUS9312388
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12388
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 486:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human T-cell Lymphotropic virus type
; INDIVIDUAL ISOLATE: III (HIV-1)
PCT-US93-12388-486

Query Match 78.7%; Score 11.8; DB 5; Length 50;
Best Local Similarity 86.7%; Pred. No. 5.8e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15
Db 24 TAAGCAGCTGCTTTT 38

RESULT 15
PCT-US93-12388-487
; Sequence 487, Application PC/TUS9312388
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
```

```
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12388
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 487:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Aids-associated retrovirus
; INDIVIDUAL ISOLATE: (arv-2;proviral)
PCT-US93-12388-487

Query Match 78.7%; Score 11.8; DB 5; Length 50;
Best Local Similarity 86.7%; Pred. No. 5.8e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15
Db 24 TAAGCAGCTGCTTTT 38

Search completed: October 2, 2001, 16:03:54
Job time: 14598 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 16:18:47 ; Search time 876.95 seconds  
(without alignments)  
10.740 Million cell updates/sec

Title: US-09-757-100B-30

Perfect score: 15

Sequence: 1 taagcagctgccatt 15

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 854978

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	15	22	AAC65562 Human focal adhesi
2	15	100.0	20	22	AAC65542 Human focal adhesi
3	14	93.3	22	20	AA336327 Sense primer used
4	12.4	82.7	27	20	AA201016 PCR primer for PGL
5	12.4	82.7	49	21	AAA64261 PCR primer for hum
6	12	80.0	29	18	AAAT96994 Human GDP-fucose p
7	12	80.0	29	18	AAAT97584 Human GDP-fucose p
8	11.8	78.7	21	22	AAAT95258 Human gene single
9	11.8	78.7	27	18	AAAT67939 Human fittl VEGF re
10	11.8	78.7	30	19	AAAT65645 HIV-1 promoter fra
11	11.8	78.7	46	14	AAQ49401 HIV-1 TATA region.

12	11.8	78.7	50	15	AAQ69736 Human T-cell lymph
13	11.8	78.7	50	15	AAQ69737 AIDS-associated re
14	11.8	78.7	50	18	AAAT64198 Human T-cell lymph
15	11.8	78.7	50	18	AAAT64199 Aids-associated re
16	11.8	78.7	50	19	AAV65660 HIV-1 TATA region
17	11.8	78.7	50	20	AAAT17486 Test sequence from
18	11.8	78.7	50	20	AAAT17487 Polymorphic fragme
19	11.6	77.3	29	21	AAA04041 IGF-I oligonucleot
20	11.4	76.0	15	22	AAAT53844 IGF-I oligonucleot
21	11.4	76.0	15	22	AAAT53845 IGF-I oligonucleot
22	11.4	76.0	15	22	AAAT53846 Human PRO1555 reve
23	11.4	76.0	18	21	AAAT58087 Primer #141 used i
24	11.4	76.0	18	22	AAAT54534 PCR primer used to
25	11.4	76.0	20	20	AAAT54538 PCR primer used to
26	11.4	76.0	20	20	AAAT52968 Human epidermal gr
27	11.4	76.0	20	20	AAAT59550 EGFR oligonucleoti
28	11.4	76.0	20	22	AAAT77117 EGFR oligonucleoti
29	11.4	76.0	20	22	AAAT77118 Native Pseudomonas
30	11.4	76.0	22	16	AAQ98175 WMV2 Mar6 oligonuc
31	11.4	76.0	22	20	AAAT7812 Human c-myc cDNA P
32	11.4	76.0	24	21	AAAT74420 Soybean chlorophyl
33	11.4	76.0	25	17	AAAT35689 RNA component of h
34	11.4	76.0	25	17	AAAT10302 Human telomerase R
35	11.4	76.0	25	20	AAAT07280 Antisense oligonuc
36	11.4	76.0	30	19	AAAT63648 Human clone 28-1 t
37	11.4	76.0	30	20	AAAT23630 Human KM-102-deriv
38	11.4	76.0	30	20	AAAT72932 KHCV cDNA fragment
39	11.4	76.0	31	14	AAAT33374 Human biallelic po
40	11.4	76.0	31	20	AAAT06130 Human IL-8 recepto
41	11.4	76.0	39	14	AAAT43806 PCR primer pRDS1 u
42	11.4	76.0	39	19	AAAT63629 M. alpina delta-6
43	11.4	76.0	39	21	AAAT09435 Sequence at juncti
44	11.4	76.0	41	11	AAAT06571 Maize polymorphic
45	11.4	76.0	41	19	AAAT51159

#### ALIGNMENTS

RESULT 1  
AAC5562  
ID AAC65562 standard; DNA; 15 BP.  
XX  
AC AAC65562;  
XX  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human focal adhesion kinase antisense sequence #28.  
XX  
XX Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.  
XX  
OS Homo sapiens.  
XX  
XX US6133031-A.  
XX  
PD 17-OCT-2000.  
XX  
XX 19-AUG-1999; 99US-0377310.  
XX  
XX 19-AUG-1999; 99US-0377310.  
XX  
XX (ISIS-) ISIS PHARM INC.  
XX  
XX Monia BP, Gaarde WA;  
XX  
XX WPI; 2001-006141/01.  
XX  
XX New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -  
XX

PS Claim 15; Column 25; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, embryonic  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

XX Sequence 15 BP; 4 A; 4 C; 3 G; 4 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
|||||

Db 1 taagcagctgccatt 15  
|||||

RESULT 2

AAC65542  
ID AAC65542 standard; DNA; 20 BP.

AC AAC65542;

DT 12-FEB-2001 (first entry)

XX Human focal adhesion kinase antisense sequence #8.

DE Human; focal adhesion kinase; FAK; signal transduction; cancer;  
KW embryonic development disorder; angiogenic disorder; wound healing;  
KW antisense; phosphorothioate; ss.

OS Homo sapiens.

PN US6133031-A.

PD 17-OCT-2000.

XX 19-AUG-1999; 99US-0377310.

XX 19-AUG-1999; 99US-0377310.

PA (ISIS-) ISIS PHARM INC.

PI Monia BP, Gaarde WA;

DR WPI; 2001-006141/01.

XX New antisense compounds for inhibiting focal adhesion kinase  
PT expression, especially useful for inhibiting retinal  
PT neovascularization, or for diagnosing and treating e.g. colon cancer -

PS Example 2; Column 23; 30pp; English.

XX The present invention describes a number of phosphorothioate antisense  
CC sequences to the human focal adhesion kinase (FAK) protein. This protein  
CC is involved in integrin-mediated signal transduction, and is implicated  
CC in cancer, particularly colon, breast and oral tumours, embryonic  
CC development disorders, angiogenic disorders and wound healing. The  
CC antisense sequences, including the one shown here, can be used in the  
CC treatment of all of these.

XX Sequence 20 BP; 5 A; 4 C; 5 G; 6 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 20;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
|||||

Db 3 taagcagctgccatt 17  
|||||

RESULT 3

AAZ01016/c  
ID AAZ01016 standard; DNA; 22 BP.

XX AC AAZ01016;

XX 21-JUL-1999 (first entry)

DE Sense primer used to amplify human FAK cDNA.

XX Human; tumorigenicity; glycosyltransferase;  
KW malignancy; brain cancer cell; protein glycosylation; glioma;  
KW meningioma; brain tumour; FAK; PCR primer; ss.

XX OS Synthetic.

XX PN WO9924584-A1.

XX PD 20-MAY-1999.

XX 12-NOV-1998; 98WO-US24224.

XX 12-NOV-1997; 97US-0969437.

XX (NEUR-) NEUROTHEAPEUTICS.

XX PI Moskal JR, Yamamoto H;

XX DR WPI; 1999-327411/27.

XX PT Altering tumorigenicity and malignancy of brain cancer cells

XX PS Example 3; Page 34; 83pp; English.

XX PCR primers AAZ01016-28 were used to amplify human FAK cDNA, in the  
CC course of the invention. The specification describes a method for  
CC altering the tumorigenicity or malignancy of brain cancer cells by  
CC changing the activity of glycosyltransferase in the cell so that  
CC glycosylation of cellular proteins is modified. The method is applied  
CC to glioma or meningioma, for prevention or treatment of brain tumours.  
CC Measuring the level of glycosyltransferase expression in brain cells  
CC is used to detect or predict their tumorigenicity.

XX SQ Sequence 22 BP; 4 A; 7 C; 5 G; 6 T; 0 other;

Query Match 93.3%; Score 14; DB 20; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 14  
|||||

Db 14 TAAGCAGCTGCCAT 1

RESULT 4

AAZ01016/c  
ID AAZ01016 standard; DNA; 27 BP.

XX AC AAZ01016;

XX 27-SEP-1999 (first entry)

DE PCR primer for PGI gene exon border.

XX PGI gene; biallelic marker; PCR primer; PGI-related biallelic marker;  
KW cancer; prostate cancer; diagnosis; therapy; prostate specific antigen;  
KW PSA; human; ss.





CC beta L-fucose (GDP-fucose) pyrophosphorylase (GDPFPP) cDNA. Cells can be  
 CC genetically engineered to contain the GDPFPP nucleic acids and produce  
 CC the protein. GDPFPP enzymes can be used to synthesise carbohydrate  
 CC molecules of defined structures, useful in investigating the role of  
 CC carbohydrates as recognition elements on cell surfaces. They are  
 CC especially useful in producing donor substrates (e.g. GDP-fucose) in  
 CC reactions of a glycosyl transferase (e.g. fucosyltransferase) with the  
 CC substrate, an acceptor sugar and a divalent metal cation to allow  
 CC formation of glycosidic linkages adding a saccharide to a substrate  
 CC saccharide. The proteins can also be used to produce antibodies or  
 CC antisera useful to characterise, detect and isolate proteins  
 CC cross-reacting with the GDPFPP protein.  
 CC Note: The specification refers to claimed GDPFPP nucleotide and  
 CC amino acid sequences, neither of which are given in the specification.  
 XX  
 SQ Sequence 29 BP; 6 A; 6 C; 10 G; 7 T; 0 other;

Query Match 80.0%; Score 12; DB 18; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 agcagctgccat 14  
 Db 27 AGCAGCTGCCAT 16

RESULT 7  
 AAT97584/C  
 ID AAT97584 standard; DNA; 29 BP.

XX AC AAT97584;

XX DT 14-APR-1998 (first entry)

XX DE Human GDP-fucose pyrophosphorylase cDNA 5' PCR primer.

XX KW GDP-fucose pyrophosphorylase; GDPFPP; human; oligosaccharide;  
 XX KW carbohydrate; PCR; primer; ss.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN W09737683-A1.

XX PD 16-OCT-1997.

XX PF 10-APR-1997; 97WO-US05968.

XX PR 09-APR-1997; 97US-0831590.

XX PR 10-APR-1996; 96US-0015241.

XX PA (CYTE-) CYTEL CORP.

XX PI Ketcham CM;

XX WPI; 1997-512415/47.

XX GDP-fucose pyrophosphorylase and related coding sequences - useful  
 PT to synthesise GDP-fucose as substrate for producing specific  
 PT carbohydrate structures, e.g. to study cell surface recognition

PS Example 2; Page 24; 40pp; English.

XX A 5' primer (AAT97584) and a 3' primer (AAT97585) were used in the PCR  
 CC amplification of human GDP-fucose pyrophosphorylase (GDPFPP) cDNA  
 CC (see AAT97583) from Epstein-Barr virus-transformed B lymphoblastoid  
 CC cell line JY, and were designed to incorporate EcoRV sites at the  
 CC 5' and 3' termini of the amplified sequence. The amplified GDPFPP  
 CC cDNA was subsequently incorporated into expression vectors.  
 CC Recombinant human GDPFPP was expressed in COS, NSO and Spodoptera  
 CC frugiperda Sf9 cells. The claimed enzyme (see AAT97581) is useful in  
 CC the synthesis of carbohydrate structures.

XX SQ Sequence 29 BP; 6 A; 6 C; 10 G; 7 T; 0 other;

Query Match 80.0%; Score 12; DB 18; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 agcagctgccat 14  
 Db 27 AGCAGCTGCCAT 16

RESULT 8  
 AAF95258/C  
 ID AAF95258 standard; DNA; 21 BP.

XX AC AAF95258;

XX DT 06-JUN-2001 (first entry)

XX DE Human gene single nucleotide polymorphism #19.

XX KW Human; variant thrombospondin 1; variant thrombospondin 4; SNP;  
 KW polymorphism; vascular disease; coronary artery disease; forensics;  
 KW myocardial infarction; atherosclerosis; stroke; venous thromboembolism;  
 KW pulmonary embolism; paternity test; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 FT Variation replace(11,T)  
 FT /\*tag= a  
 FT /standard\_name= "single nucleotide polymorphism"

XX PN W0200118250-A2.

XX PD 15-MAR-2001.

XX PF 07-SEP-2000; 2000WO-US24503.

XX PR 10-SEP-1999; 99US-0153357.

XX PR 26-JUL-2000; 2000US-0220947.

XX PR 16-AUG-2000; 2000US-0225724.

XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Lander ES, Gargill M, Ireland JS, Bolk S, Daley GQ, McCarthy JJ;

XX WPI; 2001-226749/23.

XX Nucleic acids comprising single nucleotide polymorphisms, useful in  
 PT applications such as forensics, paternity testing, medicine, genetic  
 PT analysis and phenotype correlations to diseases such as diabetes and  
 PT atherosclerosis -

PS Examples; Page 48; 242pp; English.

XX The present invention provides a method of diagnosing a vascular disease  
 CC in an individual, involving determining the sequence at various  
 CC polymorphic sites within the human thrombospondin 1 and thrombospondin 4  
 CC genes. The sequences at a number of polymorphic sites are also provided  
 CC in the specification. In particular, the method can be used in the  
 CC diagnosis of atherosclerosis, myocardial infarction, coronary heart  
 CC disease, stroke, peripheral vascular diseases, venous thromboembolism  
 CC and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also  
 CC useful in forensics, paternity testing, genetic analysis and phenotype  
 CC correlations to diseases. The present sequence is an example of one of  
 CC the human gene SNPs shown in the specification.

XX SQ Sequence 21 BP; 6 A; 7 C; 4 G; 4 T; 0 other;

Query Match 78.7%; Score 11.8; DB 22; Length 21;  
 Best Local Similarity 86.7%; Pred. No. 1.6e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 taagcagctgccatt 15  
 || |||||  
 Db 17 TAGGCAGCTGCATT 3

RESULT 9  
 AAX67939/c  
 ID AAX67939 standard; RNA; 27 BP.  
 XX  
 AC AAX67939;  
 XX  
 DT 28-JUL-1999 (first entry)  
 XX  
 DE Human flt1 VEGF receptor hammerhead ribozyme #665.  
 XX  
 KW Vascular endothelial growth factor receptor; VEGF receptor; flt-1;  
 KW flk-1; KDR; hammerhead ribozyme; hairpin ribozyme; cleavage;  
 KW tumour angiogenesis; psoriasis; rheumatoid arthritis; ocular disease;  
 KW fms-like tyrosine kinase 1; kinase insert domain containing receptor;  
 KW foetal liver kinase 1; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO9715662-A2.  
 XX  
 XX 01-MAY-1997.  
 PD  
 XX 25-OCT-1996; 96WO-US17480.  
 PF  
 XX 11-JAN-1996; 96US-0584040.  
 PR  
 XX 26-OCT-1995; 95US-0005974.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (RIBO-) RIBOZYME PHARM INC.  
 XX  
 PI Escobedo J, McSwiggen J, Pavco P, Stinchcomb D;  
 XX  
 XX WPI; 1997-259017/23.

Nucleic acid molecule modulating VEGF receptor(s) gene expression or  
 mRNA stability - useful for treating e.g. tumour angiogenesis,  
 psoriasis, rheumatoid arthritis, etc., in a human patient  
 ClaIm 9; Page 66; 218pp; English.  
 The present invention describes nucleic acid molecules which modulate  
 the synthesis, expression and/or stability of a mRNA encoding 1 or more  
 receptors of vascular endothelial growth factor (VEGF). A patient  
 (preferably human) having a condition associated with the level of the  
 fms-like tyrosine kinase 1 (flt-1), kinase insert domain containing  
 receptor (KDR) and/or foetal liver kinase 1 (flk-1) (e.g. tumour  
 angiogenesis, ocular diseases, psoriasis and rheumatoid arthritis) can  
 be treated by administering the nucleic acid molecule or the expression  
 vector to the patient. AAX67275 to AAX75752 represent specific examples  
 of nucleic acid molecules from the present invention.  
 XX  
 XX Sequence 27 BP; 9 A; 4 C; 9 G; 4 U; 1 other;

Query Match 78.7%; Score 11.8; DB 18; Length 27;  
 Best Local Similarity 86.7%; Pred. No. 1.6e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 taagcagctgccatt 15  
 || |||||  
 Db 15 TCATCAGCTGCCATT 1

RESULT 10

AAV65645  
 ID AAV65645 standard; DNA; 30 BP.

XX  
 AC AAV65645;

XX  
 DT 16-DEC-1998 (first entry)

XX  
 DE HIV-1 promoter fragment used as a target sequence.

XX  
 KW Polyamide; DNA-binding ligand; modulation; gene expression; CBP; IM;  
 KW minor groove transcription factor protein; MGTFF; TFIIA; adenocarcinoma;  
 KW transcription inhibitor; pathogen; binding affinity; zinc finger protein;  
 KW carboxamide binding pair; N-methylimidazole carboxamide; bacteria; fungi;  
 KW 5S ribosomal RNA gene; internal control region; virus; protozoa; HIV-1;  
 KW human immune deficiency virus; cancer; DNA capture; genomic sequencing;  
 KW her-2/neu oncogene; gene transcription; DNA cleavage; ds.

XX  
 OS Human immunodeficiency virus type 1.

XX  
 FH Key Location/Qualifiers  
 FT misc\_binding 6..11  
 FT FT /\*tag= a  
 FT FT /note= "six basepair sequence bound by the  
 FT FT polyamide"

FT  
 FT TATA\_signal 13..17

FT  
 FT /\*tag= b

FT  
 FT /note= "TFIID binding site"

FT  
 FT misc\_binding 18..23

FT  
 FT /\*tag= c

FT  
 FT /note= "six basepair sequence bound by the  
 FT FT polyamide"

XX  
 PN WO9835702-A1.

XX  
 PD 20-AUG-1998.

XX  
 XX 11-FEB-1998; 98WO-US02444.

XX  
 XX 21-JUL-1997; 97WO-US12722.

PR  
 PR 14-FEB-1997; 97US-0038384.

PR  
 PR 14-FEB-1997; 97US-0038394.

PR  
 PR 21-APR-1997; 97US-0853022.

XX  
 XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

PA  
 PA (SCRI ) SCRIPPS RES INST.

XX  
 PI Baird EE, Dervan PB, Gottesfeld JW, Mosier DE;

XX  
 DR WPI; 1998-506287/43.

XX  
 XX Modulating expression of genes with polyamide(s) specific for  
 PT region near the binding site for transcription factor - for  
 PT inhibiting replication of pathogen, especially human immune  
 PT deficiency virus and for treating cancers associated with expression  
 PT of the her-2/neu oncogene

XX  
 PS Example 3; Fig 17; 113pp; English.

XX  
 CC This represents a promoter fragment of the HIV-1 genome that is  
 CC used as a target sequence. The invention provides polyamide DNA-binding  
 CC ligands for modulating expression of cellular and viral genes. One method  
 CC comprises identifying a unique target DNA sequence adjacent to the  
 CC binding site of a minor groove transcription factor protein (MGTFF),  
 CC choosing a polyamide having subnanomolar affinity for the target DNA and  
 CC treating the target DNA with the polyamide as transcription inhibitor.  
 CC Methods (1) for inhibiting replication of a pathogen by administering a  
 CC transcription-inhibiting polyamide; (2) improving binding affinity of  
 CC polyamide, selected for an identified viral DNA target by replacing a  
 CC carboxamide binding pair (CBP) that does not include N-methylimidazole  
 CC carboxamide (IM) with a CBP comprising paired beta-alanine (beta)  
 CC residues; (3) inhibiting binding of the zinc finger protein TFIIA to

CC the 5S ribosomal RNA gene internal control region; (4) for treating  
 CC adenocarcinoma of ovary, endometrium, breast, fallopian tubes and cervix  
 CC with the polyamide are also provided. Method (1) is applied to viruses,  
 CC bacteria, fungi and protozoa, especially human immune deficiency virus  
 CC (HIV)-1, both therapeutically and for treating blood cells in vitro.  
 CC Method (4) is especially used against cancers that overexpress the  
 CC her-2/neu oncogene. The polyamide can also be used for diagnosis of  
 CC disease, very generally as therapeutic agents for any disease involving  
 CC cellular or viral gene transcription, for genomic sequencing, for DNA  
 CC capture and for DNA cleavage (oxidative or by light). The polyamide has  
 CC excellent specificity and very high affinity for the target DNA, which  
 CC are specific for particular genes, and are cell permeable.  
 XX  
 SQ Sequence 30 BP; 7 A; 6 C; 7 G; 10 T; 0 other;

Query Match 78.7%; Score 11.8; DB 19; Length 30;  
 Best Local Similarity 86.7%; Pred. No. 1.7e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
 ||| ||| ||| ||| |||  
 Db 15 taagcagctgctttt 29

RESULT 11  
 AAQ49401  
 ID AAQ49401 standard; CDNA; 46 BP.

AC AAQ49401;

DT 27-APR-1994 (first entry)

XX HIV-1 TATA region.

XX TATA modulating factor; TMF; transcription; TATA box; promoter; HIV-1;  
 KW human immunodeficiency virus-1; short arm; human chromosome 3; p12-p21;  
 KW translocation; cancer; ss.

OS Human immunodeficiency virus-1.

PN WO9320106-A.

PD 14-OCT-1993.

XX 31-MAR-1993; 93WO-US03077.

XX 02-APR-1992; 92US-0862025.

XX (TEXA) UNIV TEXAS SYSTEM.

PA Gaynor RB, Wu F;

PI WPI; 1993-336836/42.

DR New protein cellular factor - capable of binding double stranded  
 XX HIV-1 tata region and activating gene expression of HIV-1TR

PS Claim 3; Page 48; 75pp; English.

XX This sequence represents the TATA region of the HIV-1 LTR from -46  
 CC to -1. This region is bound by TATA modulating factor (TMF). TMF  
 CC is a protein of mol. wt. 123-130 kD which activates transcription in  
 CC most genes, esp. in human immunodeficiency virus-1 (HIV-1) by binding  
 CC to the TATA box region of the promoter. TMF is encoded by the short  
 CC arm of human chromosome 3 in the region p12-p21 which is often  
 CC involved in translocations in patients having lung and other types  
 CC of cancer.

SQ Sequence 46 BP; 8 A; 12 C; 11 G; 15 T; 0 other;

Query Match 78.7%; Score 11.8; DB 14; Length 46;

Best Local Similarity 86.7%; Pred. No. 1.7e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15  
 ||| ||| ||| ||| |||  
 Db 22 taagcagctgctttt 36

RESULT 12

AAQ69736

ID AAQ69736 standard; DNA; 50 BP.

XX AC AAQ69736;

XX 06-MAR-1995 (first entry)

XX Human T-cell lymphotropic virus type III (HIV-1), target region.

XX DNA protein-binding assay; test sequence; screening sequence;

KW promoter; target; TATA box; Herpes Simplex Virus; HSV;

KW origin of replication; UL9; transcription factor; TFIID: ds.

XX Synthetic.

XX WO9414980-A.

XX 07-JUL-1994.

XX 20-DEC-1993; 93WO-US12388.

XX 23-DEC-1992; 92US-0996783.

PR 17-SEP-1993; 93US-0123936.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;

DR WPI; 1994-234711/28.

XX Sequence-directed DNA-binding molecules - useful in  
 PT pharmaceuticals and as molecular reagents

XX Claim 28; Page 455; 587pp; English.

XX A DNA protein-binding assay is provided, useful for screening  
 CC libraries of synthetic or biological cpds. for their ability  
 CC to bind DNA test sequences. The assay is versatile in that any  
 CC number of test sequences can be tested by placing the test sequence  
 CC adjacent to a defined protein-binding screening sequence. Binding  
 CC of mols. to these test sequences changes the binding characteristics  
 CC of the protein mol. to its cognate binding sequence. When such a mol.  
 CC binds the test sequence, the equilibrium of the DNA:protein complexes  
 CC is disturbed, generating changes in the concentration of free DNA probe.  
 CC One application of this method is to eucaryotic general transcription  
 CC factors (e.g. TFIID), where the target region is typically selected  
 CC from DNA sequences adjacent to the binding site for the eucaryotic  
 CC transcription factor. Numerous exemplary test sequences are given:  
 CC the sequences in AAQ69251-731 and AAQ69850 correspond to promoter  
 CC targets (typically, TATA box-contg. sites) for human genes and the  
 CC sequences in AAQ69732-849 correspond to promoter targets for viral genes.  
 CC The test sequences may also be randomly generated. DNA:protein  
 CC interaction may be used for screening purposes, e.g. the Herpes Simplex  
 CC Virus (HSV) origin of replication and UL9 (see AAQ69851-52, AAQ69865 and  
 CC AAQ69891).

SQ Sequence 50 BP; 9 A; 14 C; 13 G; 14 T; 0 other;

Query Match 78.7%; Score 11.8; DB 15; Length 50;  
 Best Local Similarity 86.7%; Pred. No. 1.8e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgccatt 15

Db 24 taagcagctgctttt 38  
|||||

RESULT 13  
AAQ69737  
ID AAQ69737 standard; DNA; 50 BP.

XX AAQ69737;  
XX 06-MAR-1995 (first entry)  
XX  
XX AIDS-associated retrovirus (arv-2; proviral), target region.  
XX  
XX DNA protein-binding assay; test sequence; screening sequence;  
KW promoter; target; TATA box; Herpes Simplex Virus; HSV;  
KW origin of replication; UL9; transcription factor; TFIID: ds.  
XX  
XX Synthetic.

XX WO9414980-A.  
XX 07-JUL-1994.  
XX  
XX 20-DEC-1993; 93WO-US12388.  
XX  
XX 23-DEC-1992; 92US-0996783.  
XX 17-SEP-1993; 93US-0123936.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;

XX WPI; 1994-234711/28.

XX Sequence-directed DNA-binding molecules - useful in  
PT pharmaceuticals and as molecular reagents

XX Claim 28; Page 455; 587pp; English.

XX A DNA protein-binding assay is provided, useful for screening  
CC libraries of synthetic or biological cpds. for their ability  
CC to bind DNA test sequences. The assay is versatile in that any  
CC number of test sequences can be tested by placing the test sequence  
CC adjacent to a defined protein-binding screening sequence. Binding  
CC of mols. to these test sequences changes the binding characteristics  
CC of the protein mol. to its cognate binding sequence. When such a mol.  
CC binds the test sequence, the equilibrium of the DNA:protein complexes  
CC is disturbed, generating changes in the concentration of free DNA probe.  
CC One application of this method is to eucaryotic general transcription  
CC factors (e.g. TFIID), where the target region is typically selected  
CC from DNA sequences adjacent to the binding site for the eucaryotic  
CC transcription factor. Numerous exemplary test sequences are given:  
CC the sequences in AAQ69251-731 and AAQ69850 correspond to promoter  
CC targets (typically, TATA box-contg. sites) for human genes and the  
CC sequences in AAQ69732-849 correspond to promoter targets for viral  
CC genes.  
CC The test sequences may also be randomly generated. DNA:protein  
CC interaction may be used for screening purposes, e.g. the Herpes Simplex  
CC Virus (HSV) origin of replication and UL9 (see AAQ69851-52, AAQ69865 and  
CC AAQ69891).

XX Sequence 50 BP; 8 A; 13 C; 13 G; 16 T; 0 other;

Query Match 78.7%; Score 11.8; DB 15; Length 50;  
Best Local Similarity 86.7%; Pred. No. 1.8e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgctttt 15  
|||||

Db 24 taagcagctgctttt 38

RESULT 14  
AAT64198  
ID AAT64198 standard; DNA; 50 BP.

XX AAT64198;

XX 17-MAR-1997 (first entry)

XX Human T-cell lymphotropic virus type-III (HIV-1) TFIID binding site.  
XX Duplex DNA; target region; binding characteristic; DNA binding protein;  
KW TFIID; transcription factor; binding site; inhibition; enhance;  
KW cancer; inherited genetic disorder; ds.  
XX

XX Human lymphotropic virus type III.

XX US5578444-A.

XX 26-NOV-1996.

XX 27-JUN-1991; 91US-0723618.

XX 20-DEC-1993; 93US-0171389.

XX 27-JUN-1991; 91US-0723618.

XX 23-DEC-1992; 92US-0996783.

XX 17-SEP-1993; 93US-0123936.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;

XX WPI; 1997-020402/02.

XX Altering binding characteristics of DNA binding proteins to duplex  
PT DNA - by attaching specific small cpd. to target region close to the  
PT protein's binding site, useful in treatment of viral disease, cancer  
etc

XX Claim 6; Column 347-348; 264pp; English.

XX The sequences given in AAT63713-4312 represent duplex DNA's which act  
CC as target regions in the method of the invention. The method for  
CC altering the binding characteristics of a DNA-binding protein to duplex  
CC DNA comprises contacting the duplex DNA with a small molecule which  
CC binds sequence-specifically to a target region, where, when the small  
CC molecule is bound to the target region, it is adjacent to, but not  
CC overlapping by more than 4 bp, a binding site for a DNA-binding protein.  
CC The small molecule is added at a concentration effective to alter the  
CC binding of the DNA binding protein, pref. TFIID, to its binding site on  
CC the duplex DNA. The binding of the small molecule may inhibit or  
CC enhance the binding of the DNA-binding protein to its binding site. The  
CC compounds isolated using this method are potentially useful as  
CC therapeutic agents for treatment of any disease which involves a  
CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.  
CC The method is suitable for screening large biological or chemical  
CC libraries and allows determination of sequence-specific and relative  
CC affinities of known DNA-binding agents for different DNA sequences.  
CC The design of these duplex DNA's allows a single DNA:protein interaction  
CC to be used for screening sequence-specific, or preferential, DNA binding  
CC proteins that recognise almost any possible sequence (see also AAT49539-  
CC 74).

XX Sequence 50 BP; 9 A; 14 C; 13 G; 14 T; 0 other;

Query Match 78.7%; Score 11.8; DB 18; Length 50;  
Best Local Similarity 86.7%; Pred. No. 1.8e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 taagcagctgctttt 15  
|||||

Db 24 taagcagctgctttt 38

RESULT 15  
AAT64199  
ID AAT64199 standard; DNA; 50 BP.  
XX  
XX  
AC AAT64199;  
XX  
XX 17-MAR-1997 (first entry)  
XX  
XX Aids-associated retrovirus proviral arv2 TFIIID binding site.  
XX  
XX Duplex DNA; target region; binding characteristic; DNA binding protein;  
XX TFIIID; transcription factor; binding site; inhibition; enhance;  
XX cancer; inherited genetic disorder; ds.  
XX  
XX Aids-associated retrovirus.  
XX  
XX US5578444-A.  
XX  
XX 26-NOV-1996.  
XX  
XX 27-JUN-1991; 91US-0723618.  
XX  
XX 20-DEC-1993; 93US-0171389.  
XX 27-JUN-1991; 91US-0723618.  
XX 23-DEC-1992; 92US-0996783.  
XX 17-SEP-1993; 93US-0123936.  
XX  
XX (GENE-) GENELABS TECHNOLOGIES INC.  
XX  
XX Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;  
XX  
XX WPI; 1997-020402/02.  
XX  
XX Altering binding characteristics of DNA binding proteins to duplex  
XX DNA - by attaching specific small cpd. to target region close to the  
XX protein's binding site, useful in treatment of viral disease, cancer  
XX etc  
XX  
XX Claim 6; Column 347-348; 264pp; English.  
XX  
XX The sequences given in AAT63713-4312 represent duplex DNA's which act  
XX as target regions in the method of the invention. The method for  
XX altering the binding characteristics of a DNA-binding protein to duplex  
XX DNA comprises contacting the duplex DNA with a small molecule which  
XX binds sequence-specifically to a target region, where, when the small  
XX molecule is bound to the target region, it is adjacent to, but not  
XX overlapping by more than 4 bp, a binding site for a DNA-binding protein.  
XX The small molecule is added at a concentration effective to alter the  
XX binding of the DNA binding protein, pref. TFIIID, to its binding site on  
XX the duplex DNA. The binding of the small molecule may inhibit or  
XX enhance the binding of the DNA-binding protein to its binding site. The  
XX compounds isolated using this method are potentially useful as  
XX therapeutic agents for treatment of any disease which involves a  
XX specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.  
XX The method is suitable for screening large biological or chemical  
XX libraries and allows determination of sequence-specific and relative  
XX affinities of known DNA-binding agents for different DNA sequences.  
XX The design of these duplex DNA's allows a single DNA:protein interaction  
XX to be used for screening sequence-specific, or preferential, DNA binding  
XX proteins that recognise almost any possible sequence (see also AAT49539-  
XX 74).

Search completed: October 2, 2001, 16:18:48  
Job time: 15492 sec

Sequence 50 BP; 8 A; 13 C; 13 G; 16 T; 0 other;

Query Match 78.7%; Score 11.8; DB 18; Length 50;  
Best Local Similarity 86.7%; Pred. No. 1.8e+03;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 1 taagcagctgcatt 15  
|||||  
Db 24 taagcagctgctttt 38









; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-377-310-3

Query Match 100.0%; Score 15; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgggctcacagtgg 15  
|||||

Db 3 gcgggctcacagtgg 17

## RESULT 3

US-09-282-996-18/C  
; Sequence 18, Application US/09282996  
; Patent No. 6143502  
; GENERAL INFORMATION:  
; APPLICANT: Grentzmann, Guido  
; APPLICANT: Gesteland, Raymond F.  
; APPLICANT: Atkins, John F.  
; TITLE OF INVENTION: Dual-Luciferase Reporter System  
; FILE REFERENCE: T5864.NP  
; CURRENT APPLICATION NUMBER: US/09/282,996  
; CURRENT FILING DATE: 1999-03-31  
; EARLIER APPLICATION NUMBER:  
; EARLIER FILING DATE:  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: WordPerfect 8.0  
; SEQ ID NO 18  
; LENGTH: 31  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Primer f5A22  
US-09-282-996-18

Query Match 89.3%; Score 13.4; DB 3; Length 31;  
Best Local Similarity 93.3%; Pred. No. 63;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcgggctcacagtgg 15  
|||||

Db 23 GCGGGATCACAGTGG 9

## RESULT 4

US-08-732-612-6  
; Sequence 6, Application US/08732612  
; Patent No. 5922583  
; GENERAL INFORMATION:  
; APPLICANT: MORSEY, MOHAMAD A.  
; TITLE OF INVENTION: METHODS FOR PRODUCTION OF RECOMBINANT  
; TITLE OF INVENTION: PLASMIDS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 PENNSYLVANIA AVE., NW  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/732,612  
; FILING DATE: 16-OCT-1996  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 29311-20006.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030 MRSNFOERSWSH  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 31 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: prim\_transcript  
; LOCATION: 1..31  
; OTHER INFORMATION: /note= "downstream primer  
; OTHER INFORMATION: designated murF2"  
US-08-732-612-6

Query Match 78.7%; Score 11.8; DB 2; Length 31;  
Best Local Similarity 86.7%; Pred. No. 4.6e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcgggctcacagtgg 15  
|||||

Db 12 GCAGGCTGACAGTGG 26

## RESULT 5

US-08-732-612-11  
; Sequence 11, Application US/08732612  
; Patent No. 5922583  
; GENERAL INFORMATION:  
; APPLICANT: MORSEY, MOHAMAD A.  
; TITLE OF INVENTION: METHODS FOR PRODUCTION OF RECOMBINANT  
; TITLE OF INVENTION: PLASMIDS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 PENNSYLVANIA AVE., NW  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/732,612  
; FILING DATE: 16-OCT-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 29311-20006.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030 MRSNFOERSWSH  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 39 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: prim\_transcript  
; LOCATION: 1..39

; OTHER INFORMATION: /note= "downstream primer  
; OTHER INFORMATION: designated murf4"  
US-08-732-612-11

Query Match 78.7%; Score 11.8; DB 2; Length 39;  
Best Local Similarity 86.7%; Pred. No. 4.6e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcggctcacagtgg 15  
||| ||| ||| ||| |||  
Db 20 GCAGGCTGACAGTGG 34

## RESULT 6

US-09-161-015-17/c  
; Sequence 17, Application US/09161015A  
; Patent No. 5965370

; GENERAL INFORMATION:  
; APPLICANT: Lex M. Cowsett  
; TITLE OF INVENTION: ANTISENSE MODULATION OF RHO G EXPRESSION  
; FILE REFERENCE: RTS-0015  
; CURRENT APPLICATION NUMBER: US/09/161.015A  
; CURRENT FILING DATE: 1998-09-25  
; NUMBER OF SEQ ID NOS: 47

; SEQ ID NO 17  
; LENGTH: 18  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Antisense Oligonucleotide  
US-09-161-015-17

Query Match 76.0%; Score 11.4; DB 2; Length 18;  
Best Local Similarity 92.3%; Pred. No. 7.5e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 14  
||| ||| ||| ||| |||  
Db 17 CGGGCGCACAGTG 5

## RESULT 7

US-08-445-289B-25/c  
; Sequence 25, Application US/08445289B  
; Patent No. 5693467

; GENERAL INFORMATION:  
; APPLICANT: Roblin III, Richard O.  
; APPLICANT: Hu, Mendong  
; APPLICANT: Tang, Jane S.  
; APPLICANT: Lee, Sunmin  
; TITLE OF INVENTION: A Mycoplasma PCR Testing System Using A  
; TITLE OF INVENTION: Set of Mixed and Single Sequence Primers  
; Patent No. 5693467

; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Type Culture Collection  
; STREET: 12301 Parklawn Drive  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: U.S.A.  
; ZIP: 20852

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,289B  
; FILING DATE: 19-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: Bade, Annette L.  
; REGISTRATION NUMBER: 37,029  
; REFERENCE/DOCKET NUMBER: RD000003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-231-5520  
; TELEFAX: 310-816-4366

; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-445-289B-25

Query Match 76.0%; Score 11.4; DB 1; Length 22;  
Best Local Similarity 92.3%; Pred. No. 7.5e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 gggctcacagtgg 15  
||| ||| ||| ||| |||  
Db 13 GGGCACACAGTGG 1

## RESULT 8

US-08-859-998-856/c  
; Sequence 856, Application US/08859998  
; Patent No. 5994076

; GENERAL INFORMATION:  
; APPLICANT: Chenchik, Alex  
; APPLICANT: Jokhadze, George  
; APPLICANT: Bibilashvili, Robert  
; TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1375  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 2200 Sand Hill Road, Suite 100  
; CITY: Menlo Park  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94025

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/859,998  
; FILING DATE: 21-MAY-1997  
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:

; NAME: Field, Bret E.  
; REGISTRATION NUMBER: 37,620  
; REFERENCE/DOCKET NUMBER: 09096/002001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-322-5070  
; TELEFAX: 415-854-0875

; INFORMATION FOR SEQ ID NO: 856:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; OTHER INFORMATION: oligonucleotide primer  
US-08-859-998-856

Query Match 73.3%; Score 11; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 ggtcacagt 14  
Db 11 GGCTCACAGT 1

## RESULT 9

US-07-959-946-8/c  
; Sequence 8, Application US/07959946  
; Patent No. 5408038  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Richard K.  
; APPLICANT: Koduri, Raju  
; APPLICANT: Young, Stephen G.  
; APPLICANT: Witzum, Joseph L.  
; TITLE OF INVENTION: Lipoprotein Assays Using Antibodies to a  
; TITLE OF INVENTION: Pan Native Epitope and Recombinant Antigens  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker &  
; ADDRESSEE: Milnamow, Ltd.  
; STREET: 180 No. 5408038th Stetson, Suite 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/959,946  
FILING DATE: 19921008  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/901,706  
FILING DATE: 18-JUN-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312)616-5400  
TELEFAX: (312)616-5460  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-959-946-8

Query Match 73.3%; Score 11; DB 1; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ggtcacagt 15  
Db 19 GCTCACAGTG 9

## RESULT 10

US-08-333-577-8/c  
; Sequence 8, Application US/08333577  
; Patent No. 5786206  
; GENERAL INFORMATION:

; APPLICANT: Smith, Richard K.  
; APPLICANT: Koduri, Raju  
; APPLICANT: Young, Stephen G.  
; APPLICANT: Witzum, Joseph L.  
; APPLICANT: Curtiss, Linda K.  
; TITLE OF INVENTION: Lipoprotein Assays Using Antibodies to a  
; TITLE OF INVENTION: Pan Native Epitope and Recombinant Antigens  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker &  
; ADDRESSEE: Milnamow, Ltd.  
; STREET: 180 No. 5786206th Stetson, Suite 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/333,577  
FILING DATE:  
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: SCRF 234.0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312)616-5400  
TELEFAX: (312)616-5460  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-333-577-8

Query Match 73.3%; Score 11; DB 1; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ggtcacagt 15  
Db 19 GCTCACAGTG 9

## RESULT 11

PCT-US92-08634-8/c  
; Sequence 8, Application PC/TUS9208634  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Richard K.  
; APPLICANT: Koduri, Raju  
; APPLICANT: Young, Stephen G.  
; APPLICANT: Witzum, Joseph L.  
; TITLE OF INVENTION: Lipoprotein Assays Using Antibodies to a  
; TITLE OF INVENTION: Pan Native Epitope and Recombinant Antigens  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker &  
; ADDRESSEE: Milnamow, Ltd.  
; STREET: 180 North Stetson, Suite 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/08634  
; FILING DATE: 19921009  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/901,706  
; FILING DATE: 18-JUN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gamson, Edward P.  
; REGISTRATION NUMBER: 29,381  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312)616-5400  
; TELEFAX: (312)616-5460  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 31 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; PCT-US92-08634-8

Query Match 73.3%; Score 11; DB 5; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gctcacagtgg 15  
|||||  
Db 19 GCTCACAGTGG 9

RESULT 12  
US-09-166-203-9  
; Sequence 9, Application US/09166203A  
; Patent No. 5968826  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, C. Frank  
; APPLICANT: Condon, Tom P.  
; APPLICANT: Cowsett, Lex M.  
; TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION  
; FILE REFERENCE: ISPH-0323  
; CURRENT APPLICATION NUMBER: US/09/166,203A  
; CURRENT FILING DATE: 1998-10-05  
; NUMBER OF SEQ ID NOS: 60  
; SEQ ID NO 9  
; LENGTH: 18  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: antisense sequence  
US-09-166-203-9

Query Match 72.0%; Score 10.8; DB 2; Length 18;  
Best Local Similarity 85.7%; Pred. No. 1.6e+03;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcgggctcacagtg 14  
|||||  
Db 1 gcggctctcagtg 14

RESULT 13  
US-08-206-176-10/c  
; Sequence 10, Application US/08206176  
; Patent No. 563940  
; GENERAL INFORMATION:  
; APPLICANT: Garner, Ian

; APPLICANT: Dalrymple, Michael A  
; APPLICANT: Prunkard, Donna E  
; APPLICANT: Foster, Donald C  
; TITLE OF INVENTION: Production of Fibrinogen in Transgenic  
; TITLE OF INVENTION: Animals  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 4225 Roosevelt Way, N.E.  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/206,176  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, Gary E  
; REGISTRATION NUMBER: 31-648  
; REFERENCE/DOCKET NUMBER: 93-15  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-547-8080 ext 322  
; TELEFAX: 206-548-2329  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 24 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: BLGAMP4  
; US-08-206-176-10

Query Match 72.0%; Score 10.8; DB 1; Length 24;  
Best Local Similarity 85.7%; Pred. No. 1.6e+03;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cgggctcacagtgg 15  
|||||  
Db 24 CTGGCTCACAGG 11

RESULT 14  
US-08-371-001-2  
; Sequence 2, Application US/08371001  
; Patent No. 5783683  
; GENERAL INFORMATION:  
; APPLICANT: Morrison Ph.D., Richard  
; TITLE OF INVENTION: Methods and Composition for Treating  
; TITLE OF INVENTION: Tumor Cells  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gray Cary Ware & Freidenrich  
; STREET: 401 "B" Street, Suite 1700  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/371,001  
; FILING DATE: January 10, 1995

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461  
REFERENCE/DOCKET NUMBER: P00095US0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 699-3630  
TELEFAX: (619) 236-1048  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Other Nucleic Acid  
US-08-371-001-2

Query Match 72.0%; Score 10.8; DB 1; Length 24;  
Best Local Similarity 85.7%; Pred. No. 1.6e+03;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 15  
||| ||||| |||||  
Db 1 CGAGCTCACTGTGG 14

RESULT 15  
US-08-756-506-9/C  
Sequence 9, Application US/08756506  
Patent No. 5905185  
GENERAL INFORMATION:  
APPLICANT: Garner, Ian  
APPLICANT: Cottingham, Ian R.  
APPLICANT: Temperley, Simon M.  
APPLICANT: Foster, Donald C.  
APPLICANT: Sprecher, Cindy A.  
APPLICANT: Prunkard, Donna E.  
TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ZymoGenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/756,506  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Sawislak, Deborah A  
REGISTRATION NUMBER: 37,438  
REFERENCE/DOCKET NUMBER: 95-28  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6672  
TELEFAX: 206-442-6678  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-756-506-9

Query Match 72.0%; Score 10.8; DB 2; Length 24;  
Best Local Similarity 85.7%; Pred. No. 1.6e+03;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgggctcacagtgg 15  
| ||||| ||||| ||  
Db 24 CTGGCTCACAGAGG 11

Search completed: October 2, 2001, 16:03:54  
Job time: 14598 sec



**THIS PAGE BLANK (USPTO)**